Research and Development – Methodology reports from Statistics Sweden 2006:2

Stochastic Population Projections for Sweden

Michael Hartmann
Gustaf Strandell

Statistiska centralbyrån Statistics Sweden

The series entitled ”**Research and Development** – **Methodology Reports from Statistics Sweden**” presents results from research activities within Statistics Sweden. The focus of the series is on development of methods and techniques for statistics production. Contributions from all departments of Statistics Sweden are published and papers can deal with a wide variety of methodological issues.

Research and Development – Methodology reports from Statistics Sweden 2006:2

Stochastic population projections for Sweden

Michael Hartman
Gustaf Strandell

Statistiska centralbyrån
2006

Research and Development – Methodology reports from Statistics Sweden 2006:2

Stochastic population projections for Sweden

Statistics Sweden
2006

Previous publication	Quantifying the quality of macroeconomic variables.
Producer	Statistics Sweden, Research and Development Department SE-701 89 ÖREBRO + 46 19 17 60 00

Inquiries	Thomas Laitila, + 46 19 17 62 18 thomas.laitila@scb.se

When quoting material from this publication, please state the source as follows:
Source: Statistics Sweden, *Stochastic population projections for Sweden.*

Cover: Ateljén SCB

ISSN 1653-7149
ISBN 91-618-1346-X
ISBN 978-91-618-1346-9
URN:NBN:SE:SCB-2006-X103OP0602_pdf

Printed in Sweden, SCB-Tryck 2006.11

Preface

This report is about predictions of demographic variables like mortality, fertility and migration. An understanding of likely compositions of the population in the future is of importance to most companies and public authorities. For instance, changes in the demographic structure affect consumer demand and labor supply. The need for the authorities to provide with public services, like health care and schooling, changes with changes in the demographics of the population. One problem on the agenda today is the problem of an aging population. The authors of this report suggest the use of stochastic models for projections of the future population. Such an approach has several appealing features. One is the ability to provide with illustrations of the uncertainty of estimates of future population characteristics. Another is the ability to illustrate the observed, random like behavior of demographic variables over time. These features are well illustrated in the report.

Statistics Sweden, October 2006

Folke Carlsson

Anna Wilén

Contents

Abstract

The report discusses the material difference between ordinary and stochastic (time series based) population projections. While ordinary population projections only provide point estimates, stochastic projections also provide prediction intervals. It is demonstrated how time series models of mortality, fertility and net-migration spawn stochastic projections that facilitate estimation of prediction intervals for the future population as well as for its characteristics e.g., the proportion aged 65 years and over. For illustrative purposes, a stochastic projection for Sweden is made for the period 2000-2030. The Technical Section presents a basic discussion of stochastic processes and relevant time series models. In addition, the Technical Section contains three papers on mortality and fertility. The papers on mortality demonstrate how the Brass logit life table survival function is made stochastic. The paper on fertility introduces a new stochastic exponential model of age-specific fertility.

1 Introduction

The current international interest in stochastic (or probabilistic) population projections is spurred by the ongoing population debate. In recent years it has been widely debated in the EU if fertility is going to recover to its former levels. This debate concerns not only if adequate labor force will be available but also if retirement schemes can sustain themselves in the future. In response to the debate, population projections[1] have become increasingly important as foundations for learned discussion.

The universal approach to making population forecasts involves soliciting expert opinions about the future. Changes in economy, technology, medicine and within other relevant areas are considered when assessing likely future demographic scenarios. Such scenario forecasts have played important roles in the past; undoubtedly they will continue to do so in the future.

Historically, there are two major occasions on which the population projection method has played a prominent role in underpinning social, economic and political discussions about the future. In France and Germany, during the latter part of the 19th century, it was often argued that projected falling fertility eventually would lead to social, economic, political and cultural stagnation. Concerns of this nature were exacerbated by the low reproduction in France and Germany during the early part of the 20th century, and especially during the 1930s. The current population debate in the European Union, the United States and in other industrialized societies shares sentiments with this much earlier debate. During the 1930s social scientists studying population growth in Africa, Asia and South America were concerned that eventually these parts of the world would suffer consequential social, economic and political hardships because of over population. Early population projections motivated the United Nations to establish programs aiming at slowing down population growth in developing

[1] Generally, a projection involves calculating the future population in response to any given set of assumptions. A forecast is a direct attempt at calculating the future population from scenario assumptions that are thought to prevail during the forecasting period.

countries. Current projections suggest that China, some two decades from now, will have more well-educated labor force than the EU and the United States taken together (Lutz and Scherbov, 2004).

The population projection method has remained an important tool for providing social scientists, planners, administrators, legislative bodies and other interested parties with demographic insights about the future. While this situation prevails today as much as it did in the past, the methodological outlook is now undergoing revision.

This revision involves that demographers and other social scientists wish to estimate the uncertainty of population projections. Such uncertainty estimates are *important* not only to users but, in addition, import insight into the behavior of demographic processes. Fundamental to amending former methodologies is the perception that demographic variables derive from stochastic processes (see Technical Documentation for definition of stochastic process).

The theory of stochastic processes, the origins of which date back to the 1930s, is the latest major contribution to mathematical statistics, and has developed rapidly after World War II. Over time, it has come to play an outstanding role in several areas of human interest, notably in physics and communication. During the past 15 years or so, the main ideas behind stochastic processes have begun to influence demographic thinking, especially in areas of population forecasting. The reason for this is that numerical calculations regarding statistical uncertainty of phenomena that unfold with time necessarily must draw their intellectual resources from this area (see e.g., Box and Jenkins, 1979; Chatfield, 1999; Chiang, 1962; Hamilton, 1994; Kendall, 1976).

This outlook however also has a temporary downside to it. Users may have difficulties understanding statistical products that derive from complex mathematical-statistical applications; for a statistical-demographic technique to be useful, necessarily it must be communicable to users. It is generally acknowledged that it will take some time for the new methodologies to fully meet the understanding of users (see e.g., Goldstein, 2004). On the other hand, the new outlook that *uncertainty* of population forecasts is a fundamental aspect is already widespread (see e.g., Alho, 1990; Keilman, Pham, and Hetland, 2002; Lee, 1998; Long and Hollmann, 2004; Lutz and Goldstein, 2004; Statistical Office of the European Communities, 1999). The present project is the result of such considerations. It should be noted that any approach to making stochastic population projections involves simulations and customized computer programming and for these reasons takes population forecasting to a higher level. To implement the modern mode of projection, national statistical offices are likely to require supportive *in-service training programs* (see e.g., Lutz and Goldstein, 2004).

For illustrative purposes, Section 2 introduces a simple stochastic model of fertility and shows how it spawns stochastic projections, which, in turn, can be used for estimating prediction intervals. Section 3 introduces stochastic models of mortality, fertility and net-migration that are implemented in stochastic projections for Sweden, 2000-2030. This section illustrates, among other things, how prediction intervals for e.g., the total population, the yearly number of births and the proportion of elderly are estimated. Section 4 concludes the paper with a discussion. Stochastic processes and relevant time series models are discussed and illustrated in Technical Documentation. This section also contains three technical papers on mortality and fertility. The papers on mortality demonstrate how the Brass logit life table model is made stochastic. The paper on fertility introduces a new stochastic exponential model of age-specific fertility.

2 Illustrative Projections for Sweden

2.1 Baseline projections

To illustrate stochastic population projections for Sweden it is convenient to have a number of ordinary (traditional) baseline projections that involve the most commonly expected levels of future fertility, mortality and migration. Table 2.1 summarizes baseline projections, referred to as Projection Series A. The initial population in these projections is the Swedish midyear population in 2000 by single-year ages and sex.

Series A (table 2.1) is typical of scenario projections in the sense that there is no time-change in mortality, fertility and migration during the period of projection. Mortality is given by the 2000 life table for Sweden (male and female life expectancies are 77.4 and 82.0 years, respectively), net-migration is confined to four different levels (ranging from 0 to 21,000 per year), and the total fertility rate, TFR, takes on levels 1.8, 1.9 and 2.0, respectively.

Table 2.1 Projection Series A: Projections with constant fertility, mortality and net-migration during 30-year projection period

In this set of projections, fertility, mortality and net-migration are constant during the projection period, 2000-2030.	Specifications			
	Constant Fertility	Constant Survival	Constant Net-migration	Initial population Sweden 2000
	Parameters			
Projection	Survival	Mean TFR	Net-migration	Projection period
1	2000 life table	1.8	0	30
2	2000 life table	1.9	0	30
3	2000 life table	2.0	0	30
4	2000 life table	1.8	15,000	30
5	2000 life table	1.9	15,000	30
6	2000 life table	2.0	15,000	30
7	2000 life table	1.8	18,000	30
8	2000 life table	1.9	18,000	30
9	2000 life table	2.0	18,000	30
10	2000 life table	1.8	21,000	30
11	2000 life table	1.9	21,000	30
12	2000 life table	2.0	21,000	30

Series A: Projected populations by sex in 2030

Projection number	Both sexes	Males	Females	TFR	Net-migration	Gain
1	8,609,021	4,295,381	4,313,640	1.8	0	-263,273
2	8,777,651	4,382,987	4,394,664	1.9	0	-94,643
3	8,947,343	4,471,142	4,476,201	2.0	0	75,049
4	9,192,599	4,588,659	4,603,940	1.8	15,000	320,305
5	9,369,921	4,680,781	4,689,140	1.9	15,000	497,627
6	9,548,311	4,773,458	4,774,853	2.0	15,000	676,017
7	9,309,305	4,647,316	4,661,989	1.8	18,000	437,011
8	9,488,359	4,740,337	4,748,022	1.9	18,000	616,065
9	9,668,493	4,833,922	4,834,571	2.0	18,000	796,199
10	9,426,029	4,705,974	4,720,055	1.8	21,000	553,735
11	9,606,815	4,799,895	4,806,920	1.9	21,000	734,521
12	9,788,691	4,894,384	4,894,307	2.0	21,000	916,397
2000 population	8,872,294	4,386,522	4,485,772			

These scenario projections show what would happen if fertility, mortality and net-migration were to remain time invariant during a thirty-year period[2]. To begin with we only focus on the projected population in the year 2030 (at this stage, we do not discuss time-varying age and sex-distributions of the population). Table 2.1 summarizes the results for Series A.

Relative to the population in 2000, in the absence of positive net-migration, a TFR of 1.8 would lead to a slightly smaller population in 2030 (-263,000 persons). A TFR of 1.9 would lead to a smaller loss (-94,600 persons), and a TFR of 2.0 would lead to a marginal gain (75,000 persons). Hence, as we shall see, although the Swedish 2000 population has an age distribution that presently promotes growth, fertility levels below replacement will necessarily lead to population sizes 30 years into the future that are more or less the same as in 2000 (Projections 1, 2 and 3)[3].

With modest net-migration, the population would gain marginally in size (Projections 4, 5 and 6). For Projection 6 the gain is about 675,000. In Projection 9 (with a TFR of 2.0 and 18,000 net-migrants per year), the gain is almost 800,000. In Projection 12, with a TFR of 2.0 and 21,000 net-migrants per year, the gain over a thirty-year period is about 900,000. Fig. 2.1 shows the growth trajectories for Projections 1, 2 and 3 and fig. 2.2 trajectories for Projections 10, 11 and 12. Graphs show lead-time from year 2000.

---

[2] Scenario projections are sometimes referred to as *if-then* projections.

[3] It should be borne in mind that different population projection programs due to their different syntaxes and rules of rounding give results that differ, if only marginally.

Fig. 2.1 Projected populations, Series A: Projections 1, 2 and 3

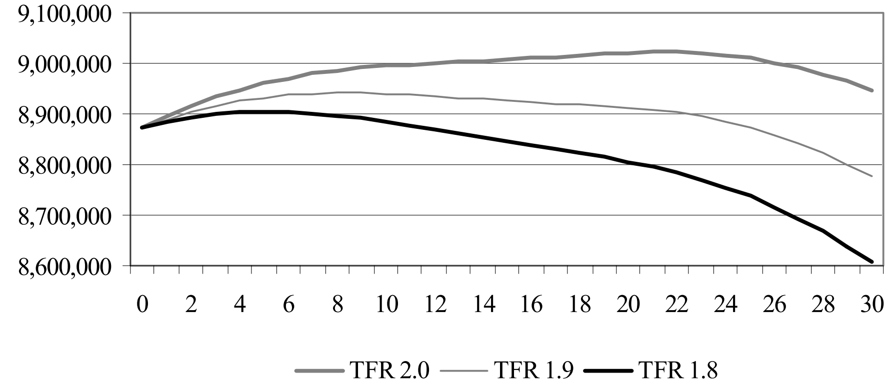
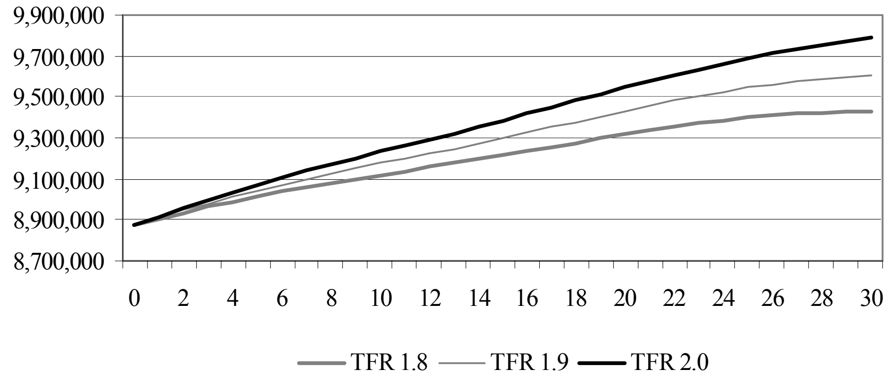


Fig. 2.2 Projected populations, Series A: Projections 10, 11 and 12



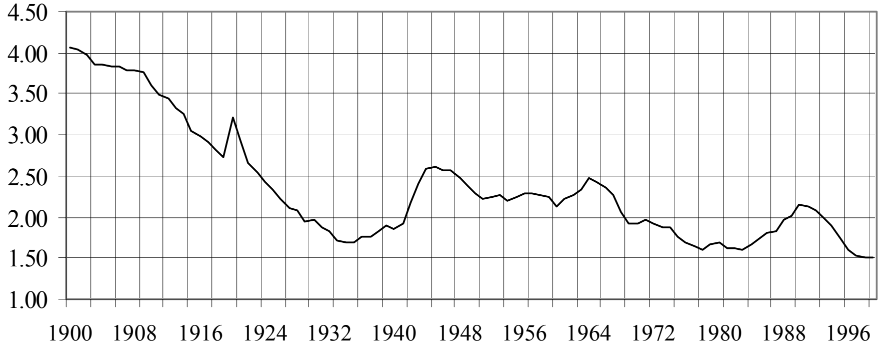
The trajectories (figs. 2.1 and 2.2) are more informative than table 2.1. Although, in the absence of positive net-migration, the population continues to grow during the early part of the projection period, eventually it begins to decline (fig. 2.1). This situation is materially changed when yearly net-migration is at the levels of Projections 4-12. However, even in the case of Projection 12, the thirty-year gain as already noted is no more than about 900,000; a yearly average increase of about 30,000 persons.

Series A (table 2.1) specifies a typical set of scenario projections. They answer simple if-then questions concerning the future. Stated differently, the answers are conditional in the sense that if assumptions are met, then the projected figures become exact forecasts.

2.2 A simple stochastic model of fertility and its implications

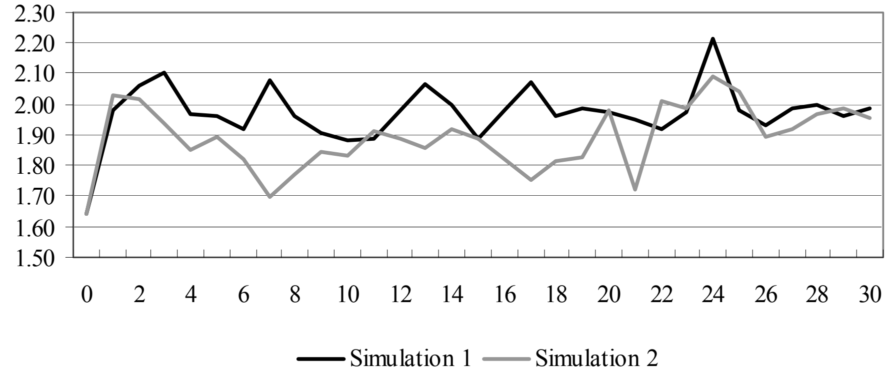
If the view is taken that any demographic variable has as its source a stochastic process, then it is obvious that the assumptions of Series A stand contradicted. After all, no live stochastic process is such that its realizations progress in time without variance. The time-pattern of TFR in Sweden for the 20th century, reproduced below from Technical Documentation, clearly indicates a process that is constantly changing, as opposed to being static year after year.

Time-pattern for total fertility rate: Sweden 1900-1999
Technical Documentation



In other words, an assumption such as TFR being constant during a long period does not belong in the demesne of stochastic processes. In fact, it could be argued that there is zero probability of any live demographic stochastic process unfolding as a straight line during any period. Hence, it also follows that any of the projections in Series A has zero probability of occurrence. It is, among other things, this statistical outlook that promotes stochastic population projections.

Fig. 2.3 TFR random walk with reflecting barriers



One of the simplest stochastic models of fertility is the random walk. Denoting TFR at time t by tfr[t], the random walk is tfr[t] = tfr[t-1] + e[t] (see Technical Documentation for details). Here reflecting barriers[4] are introduced so that

tfr[t] = tfr[t-1] + e[t]. (2.1)

with the conditions that if

tfr[t] < 1.8 then tfr[t] = 1.8 + e[t]

else, if

tfr[t] > 2.0 then tfr[t] = 2.0 + e[t]

Errors e[t] are independent and normally distributed with standard deviation $\sigma = 0.1$ (empirically determined). This means that (for practical purposes) tfr[t] is confined to moving stochastically within upper and lower boundaries. These boundaries reflect what some experts judge to be the future range of fertility in Sweden. Whenever the process runs it brings about a new time-pattern. Time patterns for two realizations of the process are shown in fig. 2.3. Despite its simplicity, the model often provides a close portrayal of time-patterns of live TFR processes (which also brings to light why it is so difficult to predict TFR).

[4] This means that if TFR falls below 1.8 then it is set at 1.8 plus an innovation. Similarly if TFR exceeds 2.0 then it is set at 2.0 plus an innovation.

Fig. 2.4 shows the result of repeating the process (2.1) 10 times, each time projecting the population 30 years into the future (with the assumptions of zero net-migration and survival as provided by the 2000 life table). This explains the computational utility of (2.1); it is a process that can be repeated in computer simulations a large number of times, each time taking the population forward any number of years. While each of the trajectories in fig. 2.4 has zero probability of repeating itself in minute detail, several realizations give us an impression of what is a likely future course of the population, subject to the validity of the specification (2.1).

Fig. 2.4. Ten projections of the Swedish 2000 both sexes population thirty years into the future assuming TFR to be a random walk

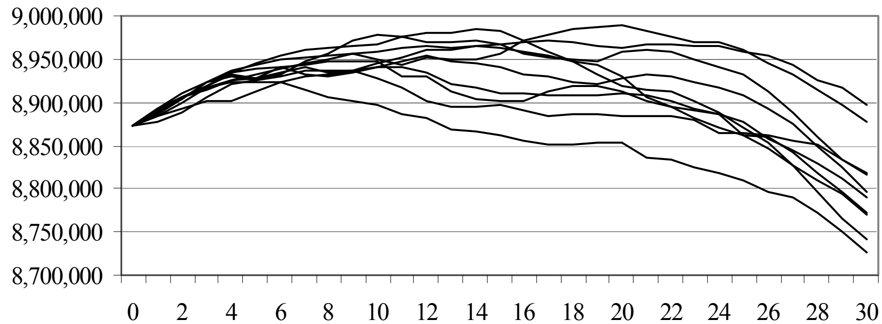


Fig. 2.5. Population mean trajectory, and estimated 95 percent confidence limits

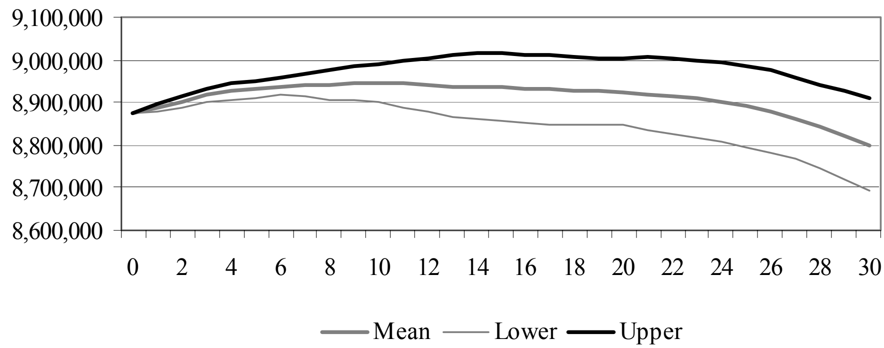


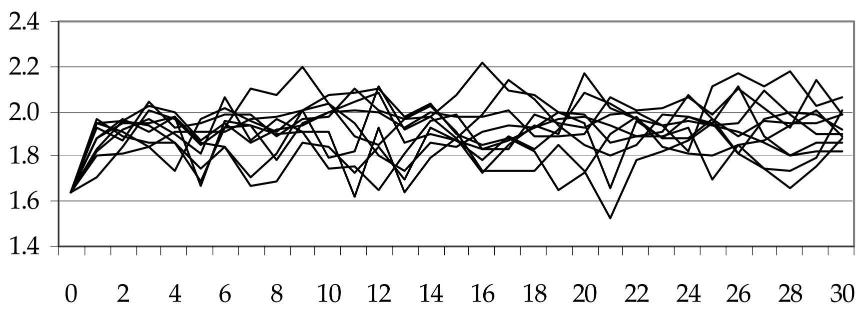
Fig. 2.4 supports three important observations. First, it will be noticed that during the early years of the projection period, the ten realizations, or time-traces, almost coincide. It is not until we are

well into the projection period that the different stochastic effects in TFR begin to actively separate one projection from another. Second, the longer we travel into the projection period, the more difference there is between trajectories, that is, *uncertainty* increases over time. Third, the projection scheme we now consider offers not just one projected population figure for each of the years during the projection period but, in fact, 10 figures. As noted, although we have only shown the result of implementing ten realizations of the simple stochastic fertility model (2.1), some very general insights are achieved!

Fig. 2.5 shows the mean of the ten simulations in fig. 2.4 together with estimated 95 percent confidence limits (which we shall also refer to as prediction intervals). It will be realized of course that fig. 2.5 serves as an illustration rather than as an attempt at accurately estimating the prediction limits (10 simulations are rather on the short side for this purpose). Fig. 2.6 shows the ten realizations of (2.1) underlying figs. 2.4 and 2.5.

As noted, the chosen fertility levels in this illustration are believed to be relevant for Sweden several years into the future. Yet, clearly there is no guarantee that these approximate fertility levels shall prevail. During the course of thirty years, numerous social, economic and world political changes may take place that influence reproductive behavior. In this sense, it is important to note that virtually any stochastic approach to making projections sooner or later must rely on subjective assessments about the future (expert panels).

Fig. 2.6. Ten realizations of TFR



3 Projections for Sweden to 2030

3.1 Time-varying fertility and net-migration: Low variation

The discussion is now extended by allowing net-migration to vary stochastically. Projection Series B assumes that fertility remains below replacement[5] level during the period 2000-2030. It is assumed that TFR will have means of 1.8, 1.9 and 2.0, respectively and follow an autoregressive AR(2) process with parameters estimated for the period 1950-2000. With respect to migration it is assumed that the net-balance will have a mean of 20,000, and follow an autoregressive AR(1) process with parameters estimated for the period 1950-2000. Hence Series B contains 3 projections with the same level of net-migration, but with varying levels of fertility (mortality is constant and given by the 2000 life table). Series B is characterized by TFR moving within rather narrow boundaries. For each level of TFR, the corresponding standard deviation of the process is $\sigma(\text{TFR}) = 0.1$. For the net-migration process, the standard deviation is $\sigma(\text{MIG}) = 5{,}000$ (table 3.1).

While in reality the net-migration process is far more complex than the simple AR(1) model, many authors favor this model not only because of its simplicity but also because of its obvious impressionistic resemblance to observed net-migration (see e.g., Congress of the United States, Congressional Budget Office, 2001; Hartmann, 2006; Sandorson, Scherbov, O'Neill and Lutz, 2004). Here it is also in place to emphasize that when stochastic models are embedded in the projection process different models may produce almost indistinguishable simulations (and prediction intervals). Fine-tuning a model of e.g., net-migration should be seen in this practical perspective.

[5] Replacement means that each woman leaves behind one daughter. To ensure this the total fertility rate must be about 2.1.

Table 3.1. Projection Series B: Projections with stochastically varying fertility and net-migration, and constant mortality during 30-year projection period: Low variation

In this set of projections, fertility and net-migration vary stochastically over time while mortality is constant during the projection period, 2000-2030.	Specifications			
	Time-varying Fertility AR(2) model $\sigma(\mathrm{TFR}) = 0.1$.	Constant Survival	Time-varying Net-migration AR(1) model $\sigma(\mathrm{MIG}) = 5{,}000$.	Initial population Sweden 2000
	Parameters			
Projection	Survival	Mean TFR	Mean net-migration	Projection period
1	2000 life table	1.8	20,000	30
2	2000 life table	1.9	20,000	30
3	2000 life table	2.0	20,000	30

Fig. 3.1. Projected mean population with lower and upper 95 percent prediction intervals. Projection 1, Series B

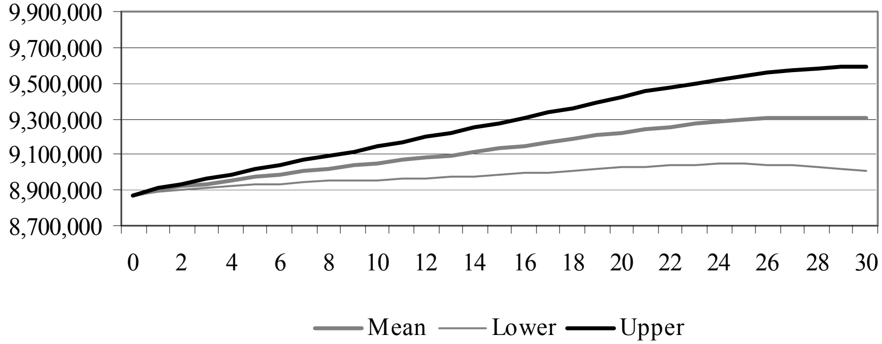


Fig. 3.2. Projected mean populations with lower and upper 95 percent prediction intervals. Projection 3, Series B

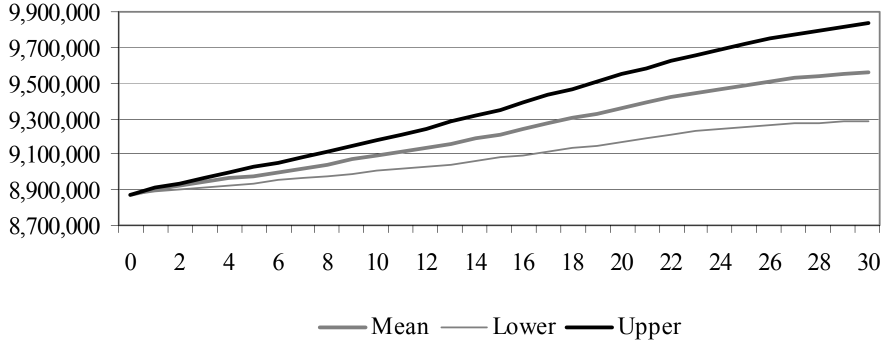


Fig. 3.3. Projected mean number of births, lower and upper 95 percent prediction intervals, and one realization of birth process. Series B, Projection 2

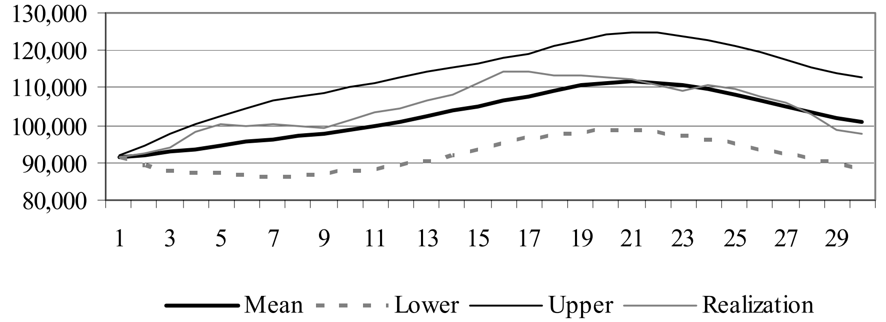


Fig. 3.4. Mean TFR, lower and upper 95 percent prediction intervals, and one realization of TFR process. Series B, Projection 3

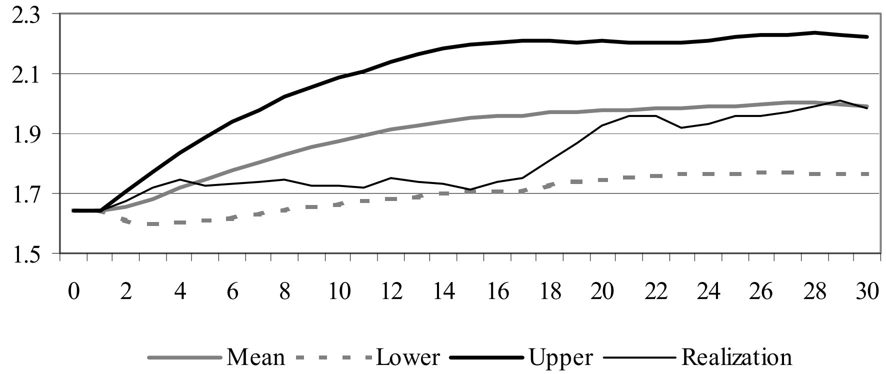
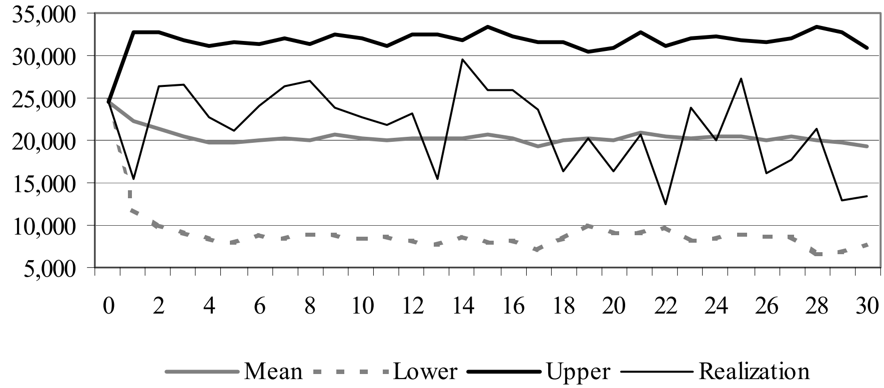


Fig. 3.5. Mean net-migration, lower and upper 95 percent prediction intervals, and one realization of net-migration process. Series B, Projection 3



Series B is a low-variation option that has been shown for two reasons. First, as noted, some experts share the view that there will be little variation in reproductive levels in the future. It is generally believed that in most parts of the EU fertility levels will remain below replacement level. While clearly it is acknowledged that within each of the member countries there will be temporal variation in period estimates of TFR, at the same time it is believed that variations will be small. Second, the stochastic population approach is an important diagnostic tool for understanding how the population and its characteristics respond to different demographic scenarios. To the extent that one has limited confidence in the above future assessments, at least Series B provides a diagnostic or hypothetical answer.

Fig. 3.1 shows the mean (150 simulations) for each year during the projection period together with upper and lower 95 percent prediction intervals for Projection 1 (Mean TFR = 1.8). It will be seen (fig. 3.1) that in 2030 the total population would range between about 9.0 and 9.6 million, with an expected mean of about 9.3 million.

Fig. 3.2 shows the corresponding results for Projection 3. With TFR = 2.0, the population in 2030 would range between 9.3 and 9.8 million, with an expected mean of about 9.6 million.

Fig. 3.3 shows for Projection 2 the mean number of births each year during the period of projection together with estimated lower and upper 95 percent prediction intervals. This time, for additional illustration, one of the birth realizations has been shown. It falls between upper and lower prediction intervals.

Fig. 3.4 shows the TFR process for Projection 3 along with one of the realizations, which, too, falls between estimated 95 percent prediction intervals.

Fig. 3.5 shows the mean number of net-migrants per year together with lower and upper 95 percent prediction intervals. For illustration one of the net-migration simulations has been shown, it falls between prediction boundaries.

3.2 Time-varying fertility and net-migration: Increased variation

Time variation in fertility and net-migration in Series B is quite small, in fact, smaller than what has been observed in the past. Series C introduces additional variation in fertility and net-migration while, as before, mortality is constant (2000 life table). The mean levels of fertility are TFRs at 1.8, 1.9 and 2.0 with $\sigma(\mathrm{TFR}) =$ 0.2. Yearly net-migration is 25,000 with $\sigma(\mathrm{MIG}) = 8{,}000$. Results are based on 150 simulations.

Fig. 3.6 shows the projected mean population (both sexes) with 95 percent prediction intervals (Series C, Projection 1). For illustration, one of the simulations is shown in fig. 3.6.

Table 3.3. Projection Series C: Projections with time-varying fertility and net-migration, and constant mortality during 30-year projection period: Increased variation

In this set of projections, fertility and net-migration vary over time while mortality is constant during the projection period, 2000-2030.	Specifications			
	Time-varying Fertility AR(2) model $\sigma(\mathrm{TFR}) = 0.2$.	Constant Survival	Time-varying Net-migration AR(1) model $\sigma(\mathrm{MIG}) =$ 8,000.	Initial population Sweden 2000
	Parameters			
Projection	Survival	Mean TFR	Mean net-migration	Projection period
1	2000 life table	1.8	25,000	30
2	2000 life table	1.9	25,000	30
3	2000 life table	2.0	25,000	30

The standard deviations for the projected populations increase to about 250,000 toward the end of the projection period (fig. 3.7). In order to receive a visual understanding of how much trajectories differ form one another, fig. 3.8 shows 25 simulations of Projection 1, Series C. A diagram of this kind is of considerable service because it shows, directly to the eye, how much dispersion there is in the process. Moreover, such a simple diagram may also indicate how many simulations are required for receiving a trustworthy image of the situation.

Fig. 3.9 shows the projected mean number of births, 95 percent prediction intervals as well as one birth realization (Series C, Projection 1).

Fig. 3.10 outlines the projected mean number of net-migrants, 95 percent prediction intervals, and one realization of net-migration (Series C, Projection 1).

Fig. 3.6. Projected mean population, 95 percent prediction intervals, and one realization. Series C, Projection 1

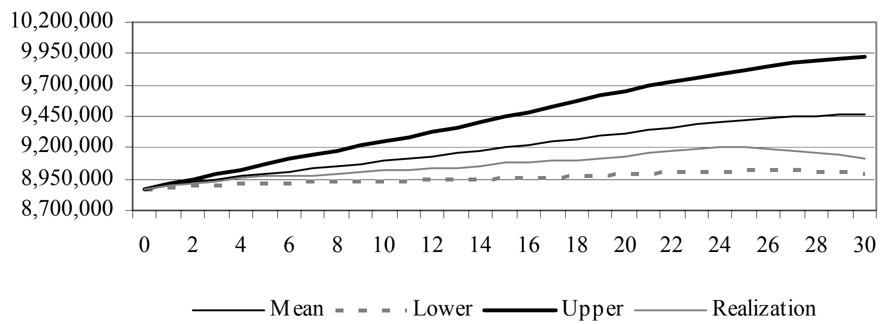


Fig. 3.7. Estimated standard deviations for projected population by year during projection period. Series C, Projection 1

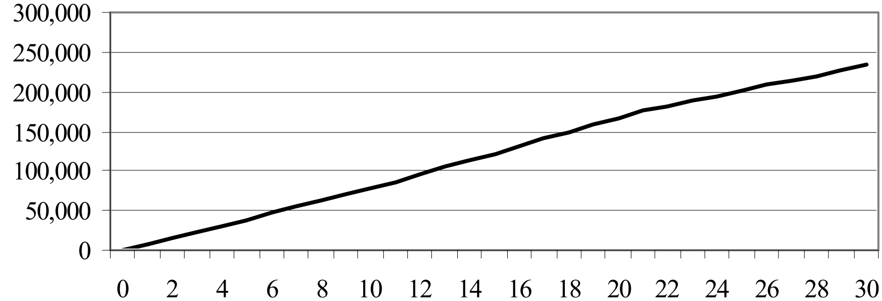


Fig. 3.8. Twenty-five realizations of Projection 1, Series C

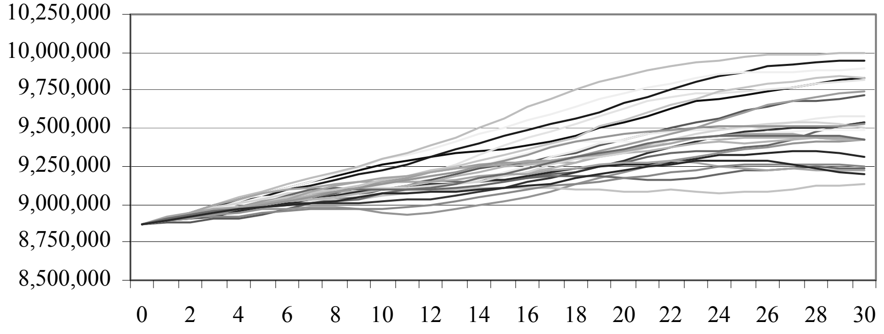


Fig. 3.9 Projected mean number of births, 95 percent prediction intervals, and one realization. Series C, Projection 1

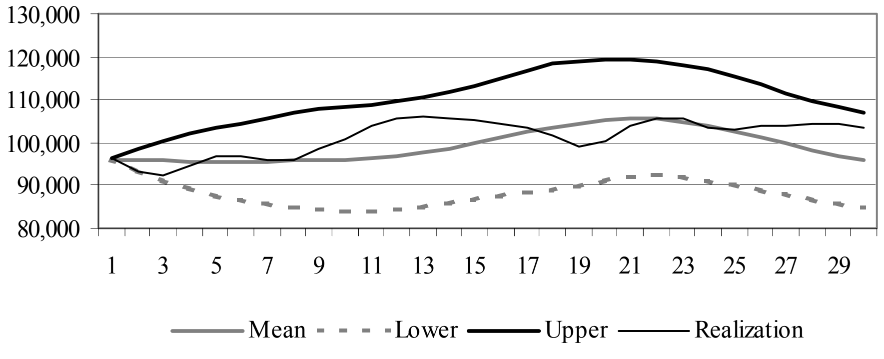
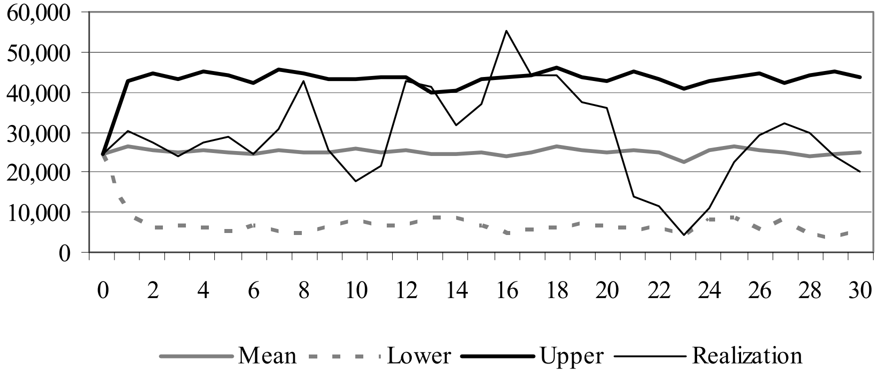


Fig. 3.10. Mean net-migration, 95 percent prediction intervals, and one realization. Series C, Projection 1



As the projection moves into the projection period, there is steadily increasing uncertainty. As noted, it is the purpose of simulation techniques to mimic this uncertainty as realistically as possible. Prediction intervals of type $\hat{\mu} \pm 2\hat{\sigma}$ however are occasionally deemed to be unrealistically wide. For this reason it is also customary to present narrower prediction intervals of type $\hat{\mu} \pm \hat{\sigma}$. Fig. 3.11 and fig. 3.12 illustrate the difference between the two types of prediction intervals (Series C, Projection 2). In both figures, 10 realizations have been included for illustration. It will be noted that relative to the 10 simulations, the prediction intervals in fig. 3.11 appear excessively wide. This is amended by fig. 3.12 with prediction intervals of type $\hat{\mu} \pm \hat{\sigma}$. In fig. 3.12 the more extreme simulations fall outside the (one-standard deviation) prediction boundaries. Fig. 3.13 shows mean number of births for Projections 1, 2 and 3 in Series C.

Fig. 3.11. Projected mean populations, two-sd prediction intrvals, and 10 realizations. Series C, Projection 2

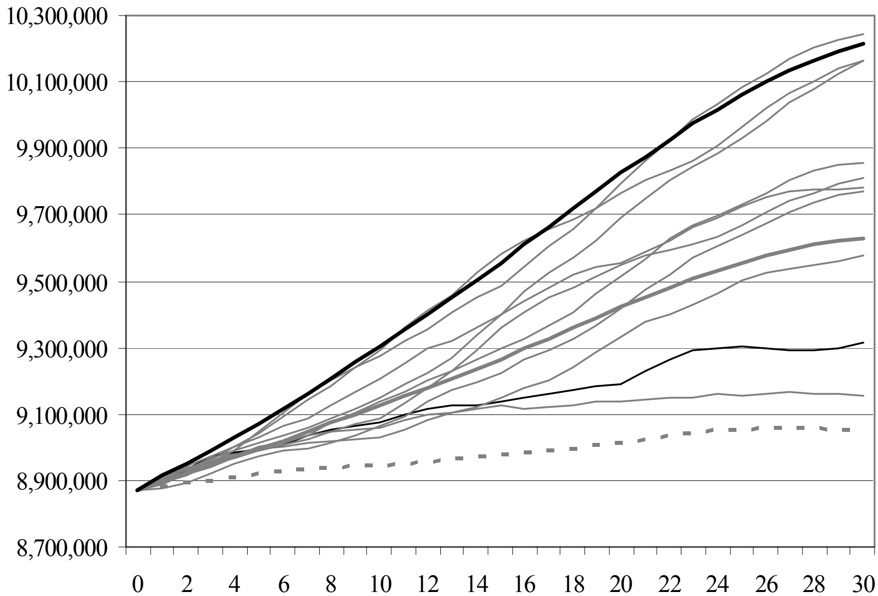


Fig. 3.12. Projected mean populations, one-sd prediction intrvals, and 10 realizations. Series C, Projection 2

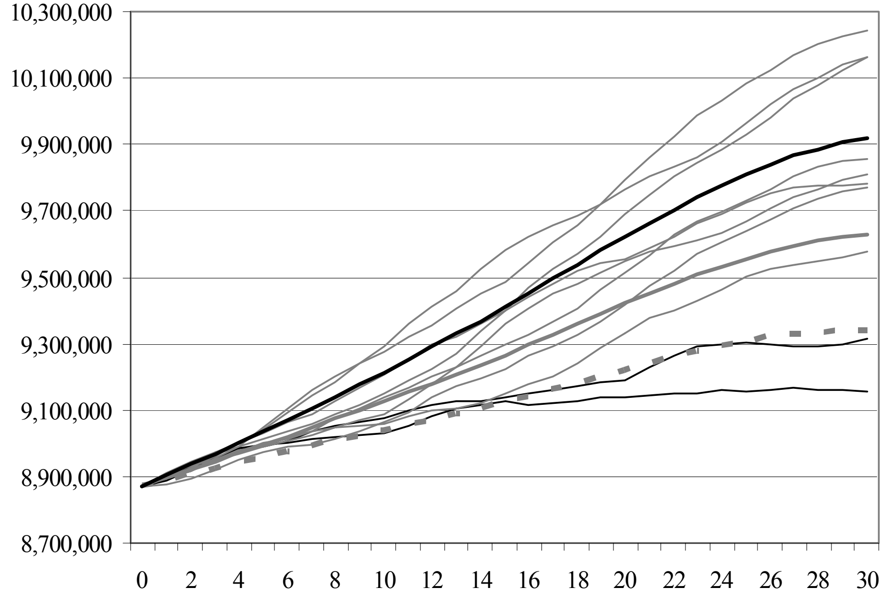
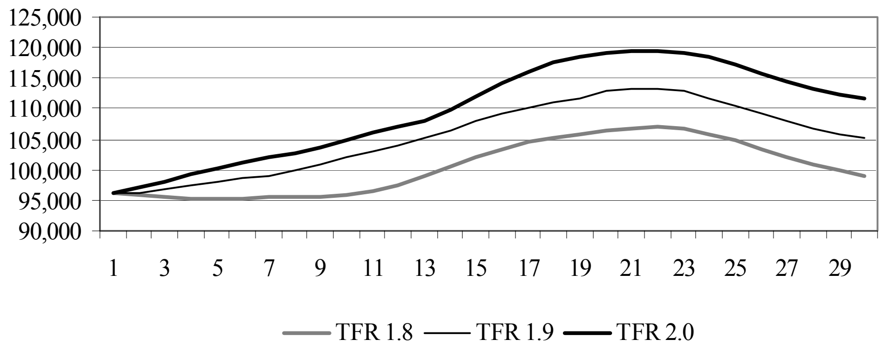


Fig. 3.13. Mean number of births per year for Projections 1, 2 and 3 Series C)



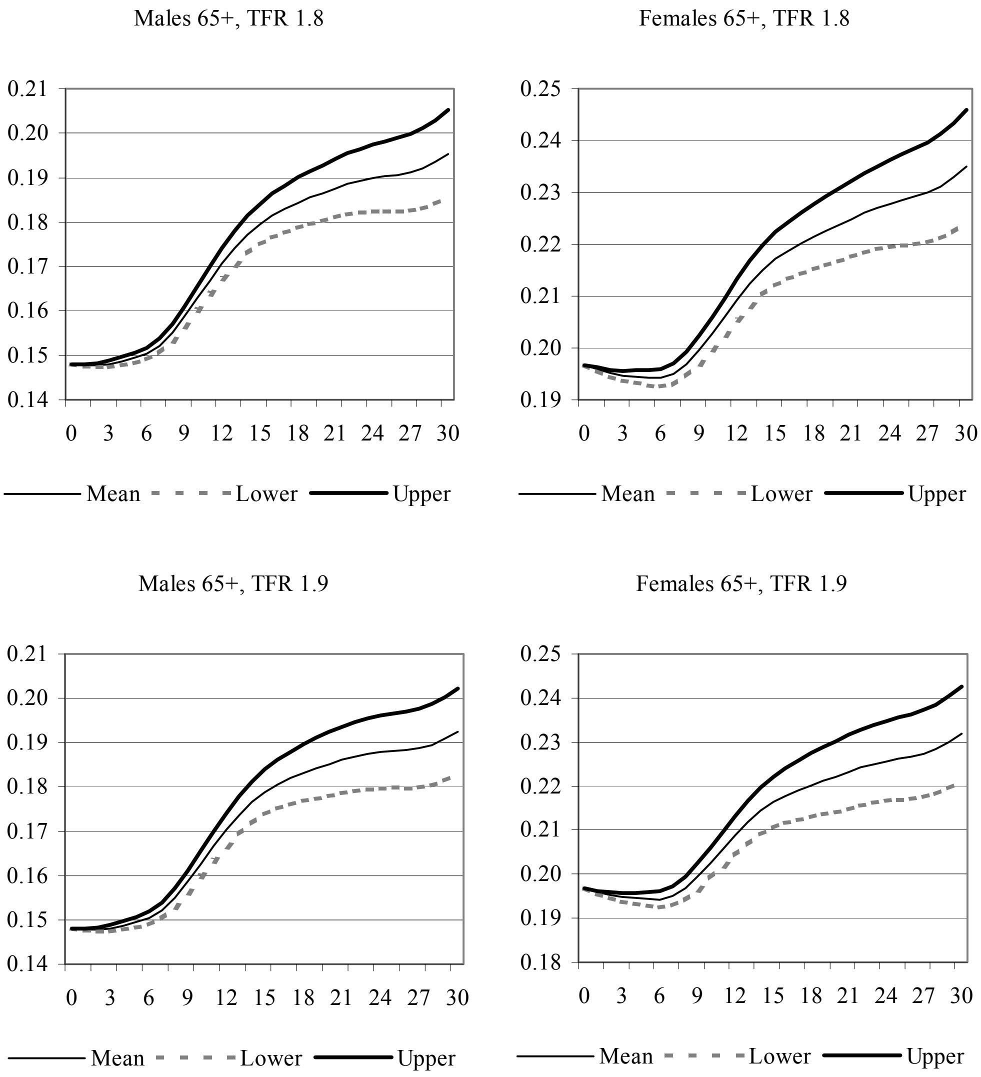
3.3 Changes in age composition

Both fertility and migration affect the age-distribution. Generally speaking, it is changing fertility that influences the age-distribution the most. Because TFR in Series C (table 3.3) runs below replacement level, it leads to an increase in the proportion of elderly persons. This increase, to some extent, is offset by net-migration.

Fig. 3.14 shows the proportions aged 65+ in Series C. During the first five years or so, there is not much change however after about 2005 the proportion aged 65+ increases steadily. For women it might reach about 25 percent in 2030 (Projection 1, Series C). In Projection 2, where the mean of TFR is 1.9, the proportion aged 65+ in 2030 could reach about 19 percent for males and 23 percent for females. Fig. 3.14 also gives 95 percent prediction intervals (150 simulations). The proportion aged 75+ in 2030 are of magnitude 9 and 12 percent for males and females, respectively (Projection 1, Series C).

From a substantive point of view it should be noted that increases in the proportion elderly also carry with them increases in health care expenditures, retirement benefits and other associated expenses. These expenditures are both personal and public (tax-financed services). Proper planning requires sound understanding of how many elderly will be part of future society. The above figures suggest that the *burden of the elderly* is due to reach high levels around 2020 and, from then on, continue to historically unprecedented levels. It is important to bear in mind however that the highest standards of living in the history of mankind were recorded during the 20th century while, at the same time, the world population and life expectancies rose to the highest levels ever recorded. In addition it must be noted that in e.g., Sweden the proportion of the population aged 65 and over has steadily increased during the 20th century; an aging population is not a new phenomenon.

Fig. 3.14. Proportions aged 65 and over, and 95 percent prediction intervals. Projections 1, 2 and 3, Series C



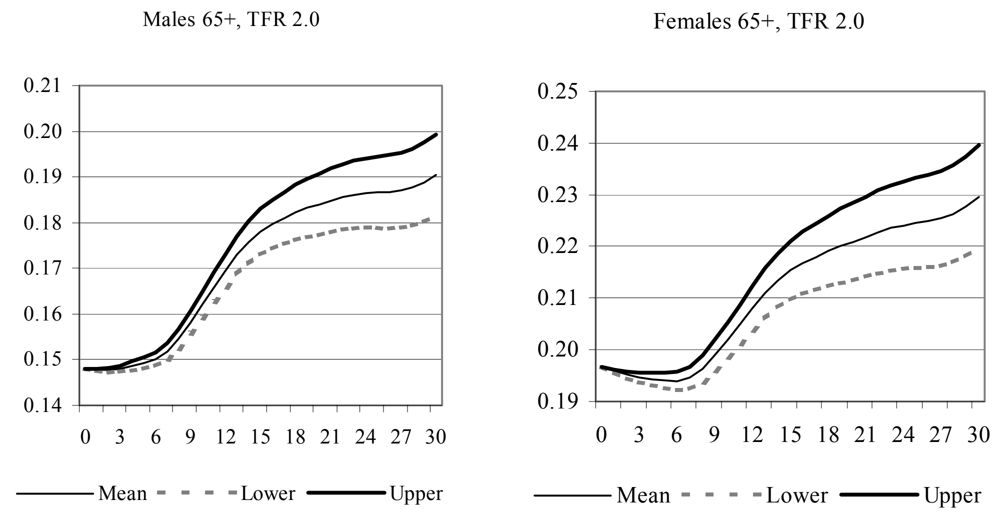
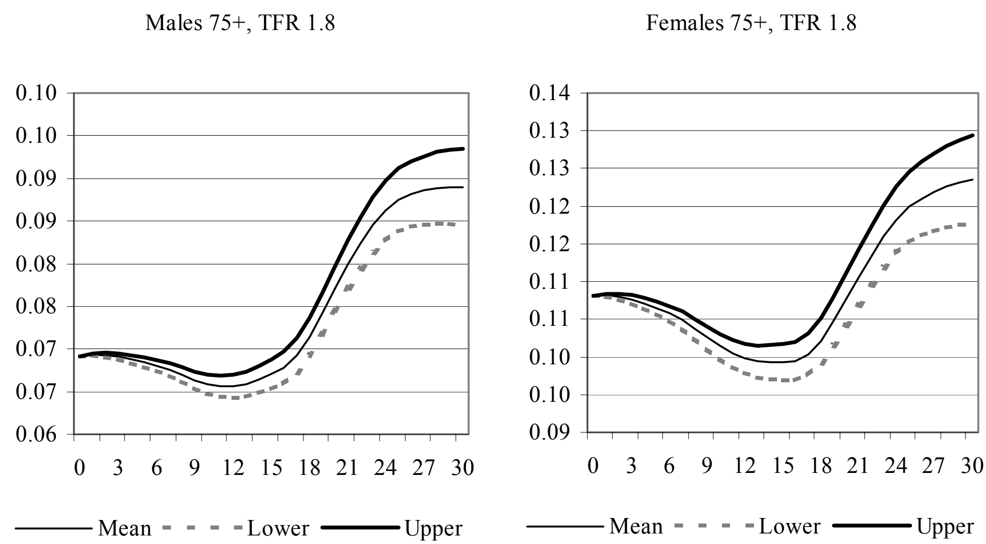


Fig. 3.15 Proportions aged 75 and over, and 95 percent prediction intervals. Projection 1, Series C



3.4 Time-varying fertility, mortality and net-migration

It is mortality, fertility and net-migration that determine the current and future population. Undoubtedly, these processes are not independent. Models that involve mutual interaction between demographic processes, and which are suitable for numerical application, are however yet to be developed. Here, as in previous examples, the assumption is that the processes are independent of one another. It is an important advantage of this approach that assessments of uncertainty can be made with respect to each individual process, mortality, fertility or net-migration (as illustrated throughout the report). In previous examples, mortality has been held constant. We now introduce variability due to mortality as well.

Table 3.6 outlines six projections (Projection Series D). In projections 1, 2 and 3 it is assumed that mortality improvements can take place at all ages (mortality model 1). At some ages improvements are miniscule. The mortality model is discussed in Technical Documentation.

In projections 4, 5 and 6 (mortality model 2) it is assumed that mortality improvements are limited to ages 55 and over while mortality below this threshold age is given by the 2000 life table. This model was developed in response to discussions with e.g., the Swedish National Insurance Board (Försäkringskassan). In addition, it is assumed that net-migration will be about 25,000 persons per year. The parameters of dispersion are given in table 3.6.

Table 3.6. Projection Series D: Projections with time-varying mortality, fertility and net-migration, during 30-year projection period

In this set of projections, mortality, fertility and net-migration vary over time during the projection period, 2000-2030.	Specifications			
	Time-varying Fertility AR(2) model σ(TFR) = 0.2.	Time-varying survival	Time-varying Net-migration AR(1) model σ(MIG) = 8,000.	Initial population Sweden 2000
	Parameters			
Projection	Survival	Mean TFR	Mean net-migration	Projection period
1	Model 1	1.8	25,000	30
2	Model 1	1.9	25,000	30
3	Model 1	2.0	25,000	30
4	Model 2	1.8	25,000	30
5	Model 2	1.9	25,000	30
6	Model 2	2.0	25,000	30

Fig. 3.16 illustrates the features of Projection 2, Series D. To the extent that TFR recovers to a level of 1.9 and provided net-migration is around 25,000 persons per year, we would expect a total popula-

tion in 2030 between 9.3 and 10.4 million. The number of births in 2030 would range between 80,000 and 130,000 (fig. 3.16).

Fig. 3.17 shows results for Projection 3, Series D. Here it is assumed that TFR recovers to a level of 2.0. This only has a marginal effect on the projected population in 2030. The 95 percent prediction intervals are now 9.4 and 10.5 million. Fig. 3.17 also illustrates the prediction intervals for TFR and one realization of TFR. Net-migration is illustrated in the same manner. During the projection period, the crude death rate, CDR, fluctuates between 9.7 and 11.4 per 1,000 population (fig. 3.17).

Fig. 3.18 shows results for Projection 4, Series D. Here mortality model 2 is invoked. The prediction intervals are a littler wider than for mortality model 1.

Fig. 3.19 shows how the age-distribution changes during the course of the projection period (Projection 4, Series D). Because the population is projected 30 years into the future, those aged 30 and over in the year 2000 are moved up in the age-distribution. Those aged below 30 in the year 2030 are the survivors of those born during the projection period supplemented by net-migrants.

We would expect a rather flat age-distribution among those aged below 30, approximately 530,000 (in 2030). At the same time, however, we would also expect uncertainty as provided by the 95 percent prediction intervals (fig. 3.19). It will be noted that by raising mean TFR from 1.8 to 1.9 only has a small effect (Projection 5, Series D). Projection 6 is the most optimistic one. Here TFR is expected to recover to a level of 2.1. The panel to the right is Projection 6 with zero net-migration, which illustrates the importance of net-migration with respect to the magnitude and age-distribution of young labor force (in 2030).

Mortality models 1 and 2 are fully explained in the Technical Documentation section. Nevertheless, a few comments are in place. Since the 1970s there has been a steady improvement in survival for both males and females. Extrapolation of these trends is difficult simply because they point to fast drops in mortality. Clearly, drops of this nature cannot for obvious reasons continue forever. Sooner or later there will be a slowing down in mortality improvement. In the case of mortality model 1 it is assumed that life expectancies will increase from about 77 to 80 years for males and from about 80 to 85 years for females between 2000 and 2030 (fig. 3.20). These increases reflect that past improvements will continue at the same pace as during the past 30 years.

Life expectancies at age 55 are also shown in fig. 3.20. For males these increase from about 25 to 27 years and for females from about 28 to 30 years. Realizations of the mortality process are shown for females (life expectancy at birth) and males (life expectancy at age 55). Fig. 3.21 shows similar graphs for mortality model 2. Fig. 3.21 compares the two mortality models.

Fig. 3.16. Both sexes, males, females, births and 95 percent prediction intervals during projection period, Projection 2, Series D

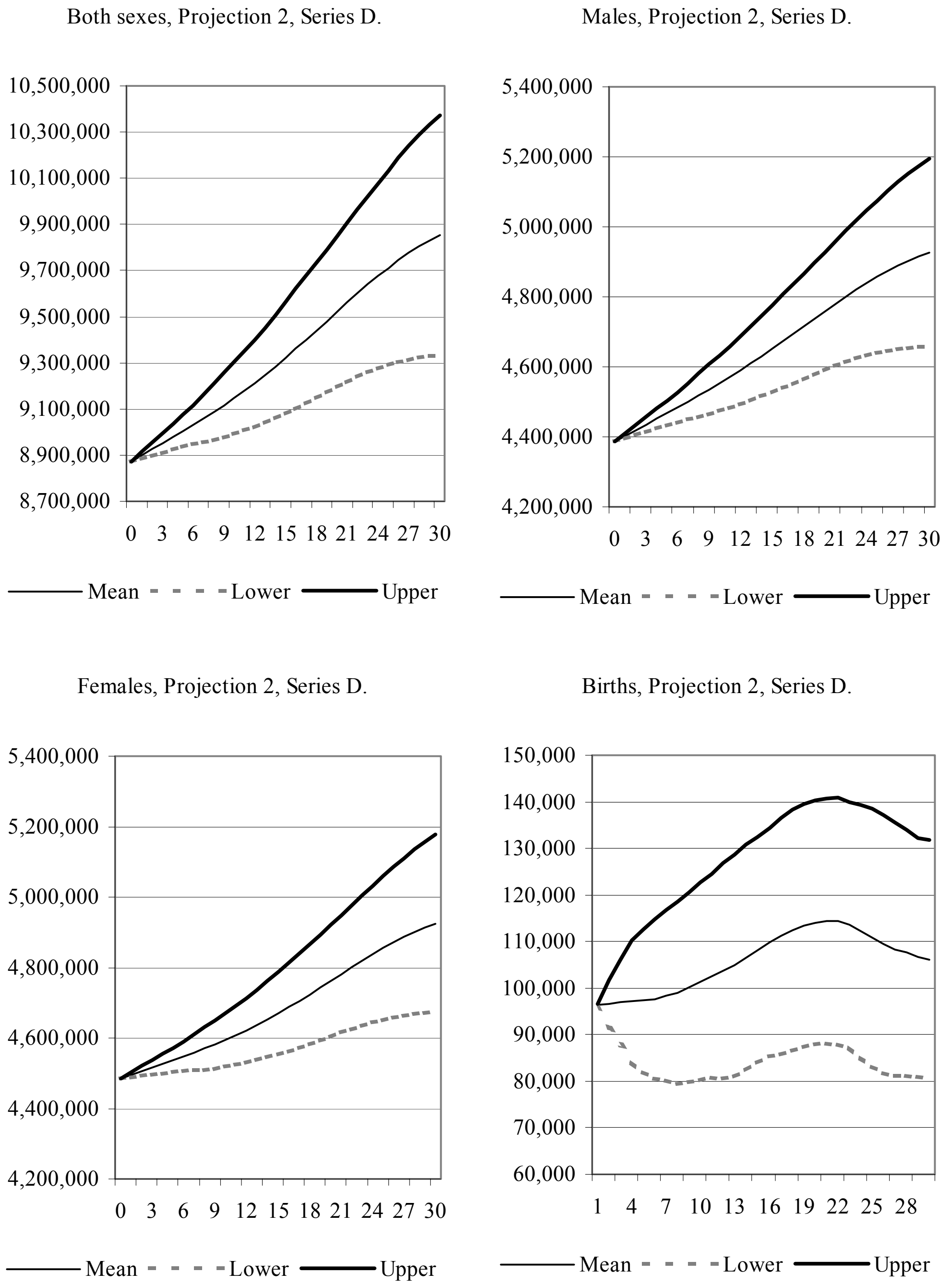
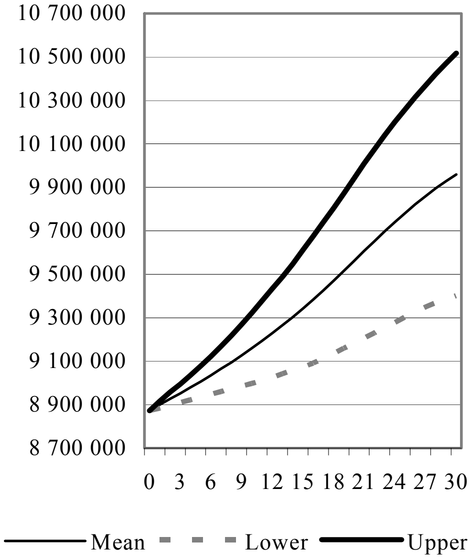
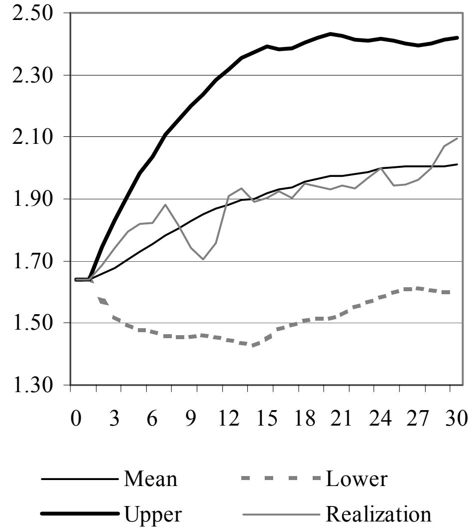


Fig. 3.17. Both sexes, TFR, net-migration and crude death rates for Projection 3, Series D

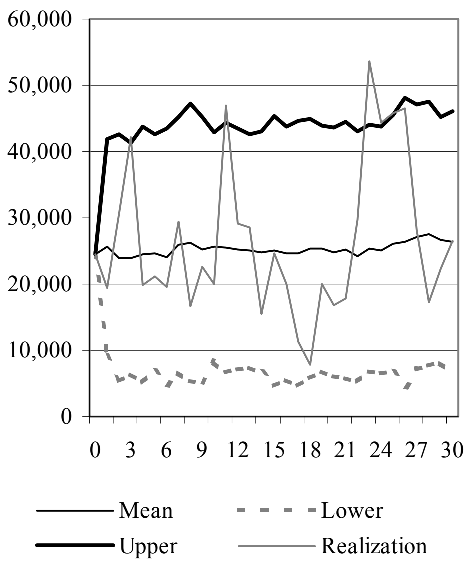
Both sexes, Projection 3, Series D.



Mean TFR and one realization, Projection 3, Series D.



Net-migration, Projection 3, Series D.



25 realizations of crude death rate, Projection 3, Series D (per 1,000 population).

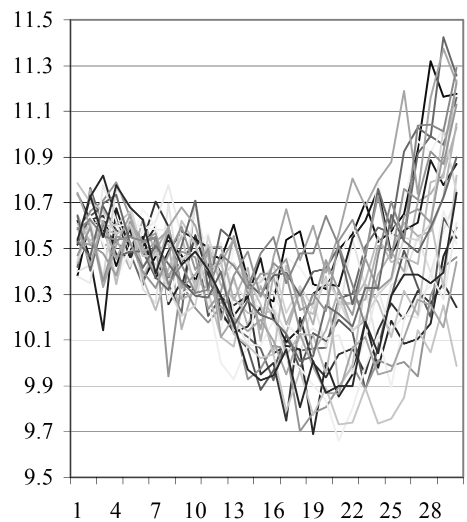


Fig. 3.18. Both sexes, males, females and births for Projection 4, Series D (mortality model 2)

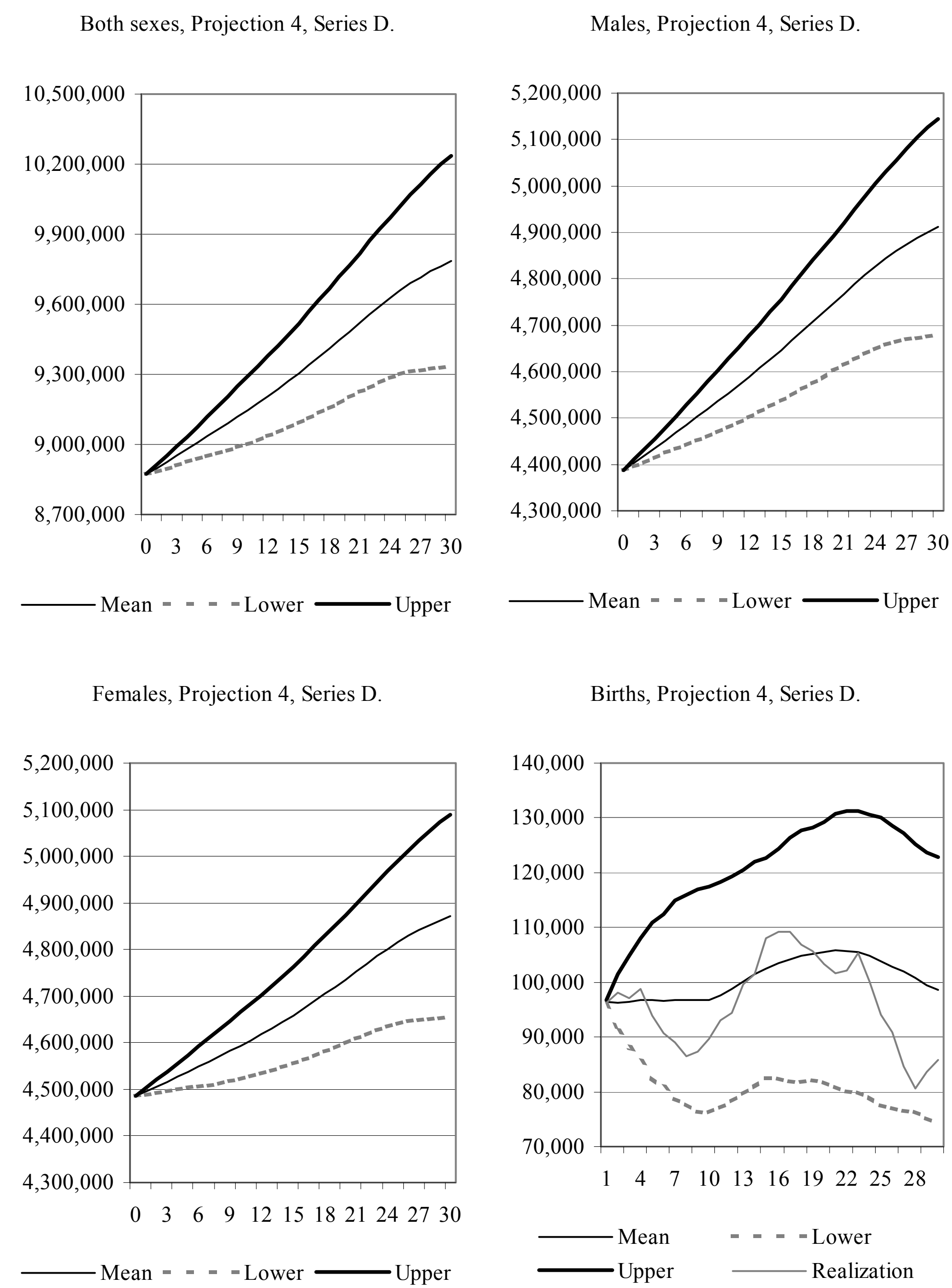


Fig. 3.19 Projected age-distributions in 2030, Projection 4, Series D with additional illustration of zero net-migration

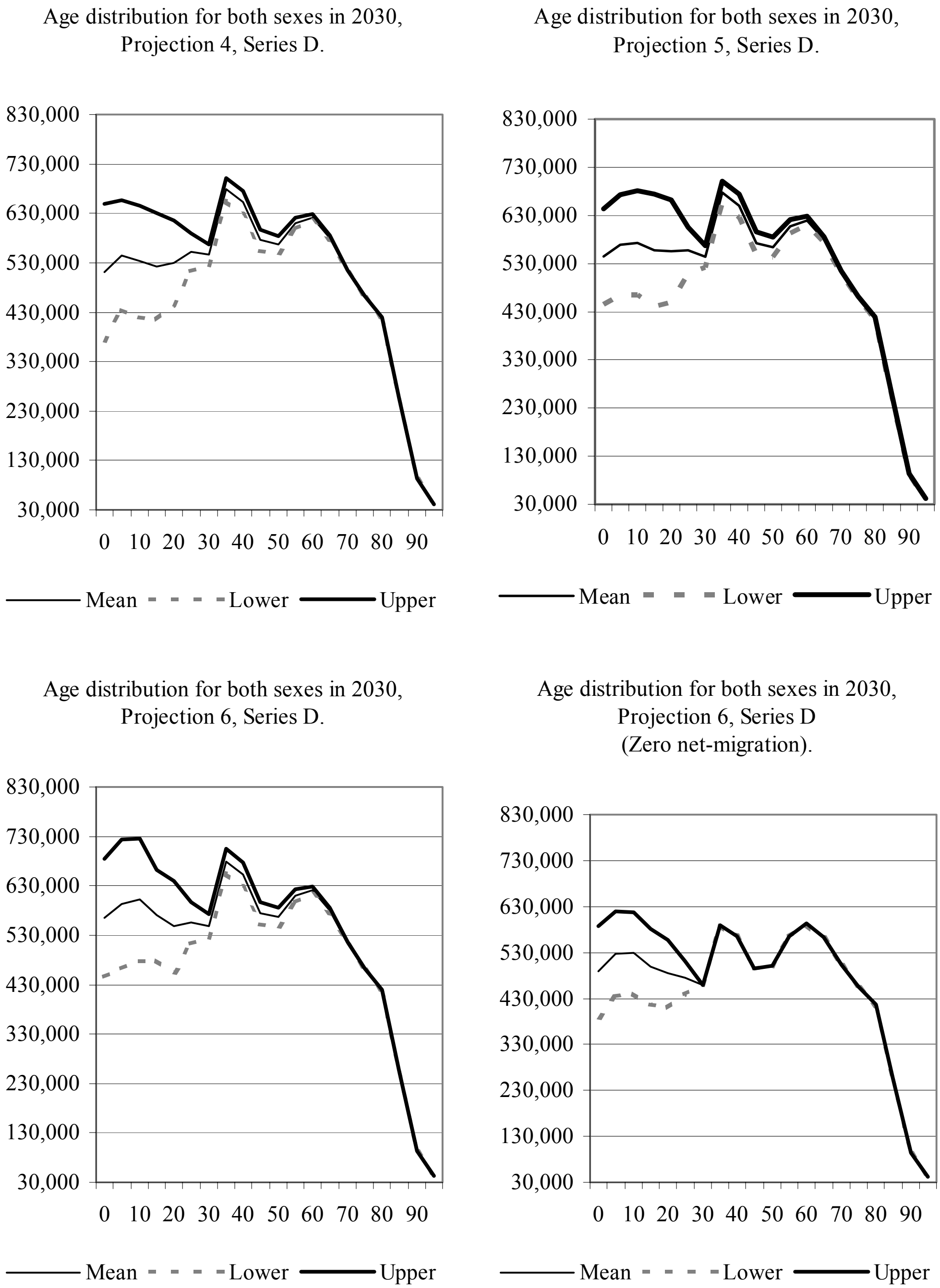


Fig. 3.20 Life expectancies at birth and at age 55, and variation in mortality model 1

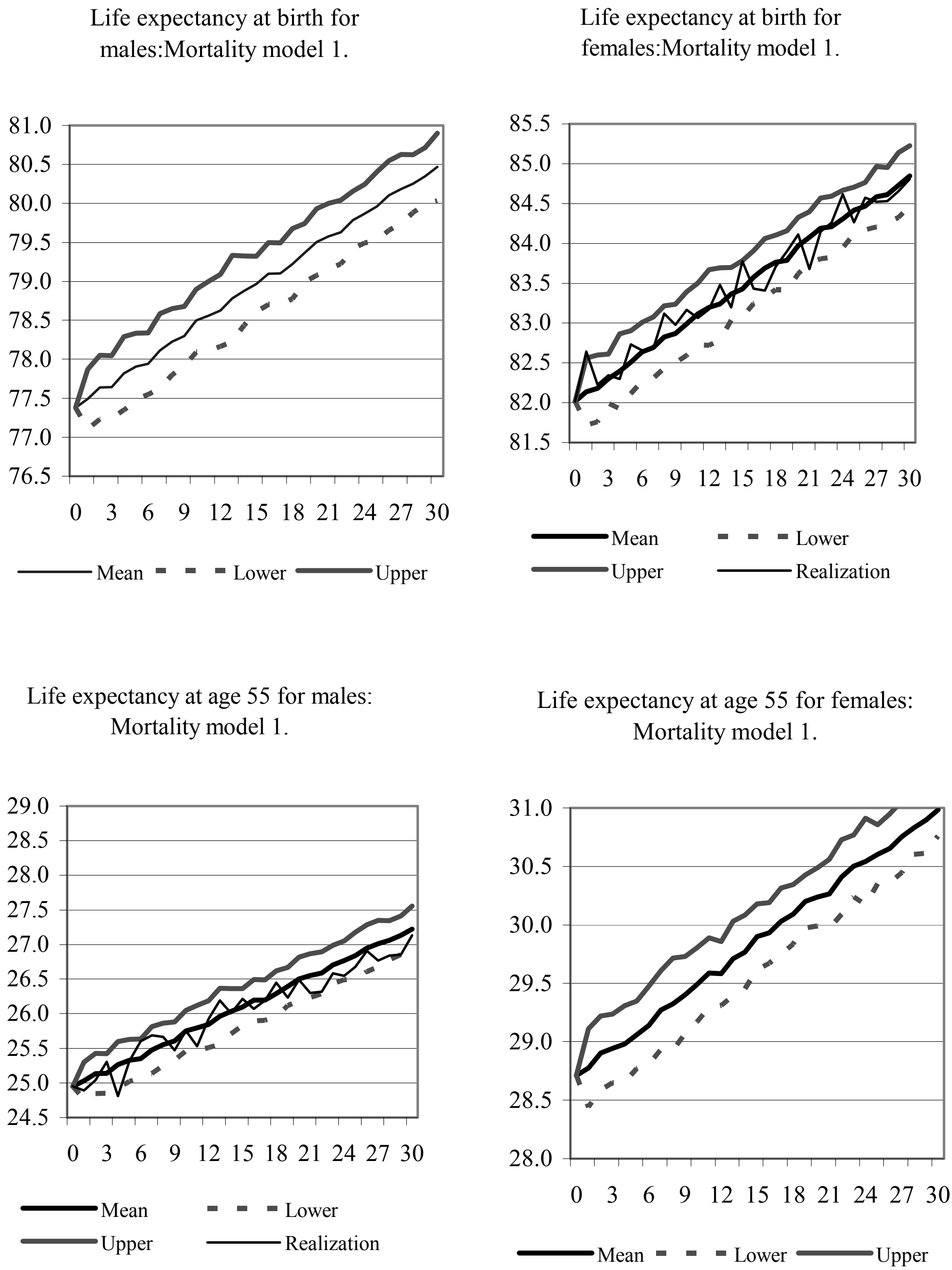


Fig. 3.21 Life expectancies and variation in mortality model 2

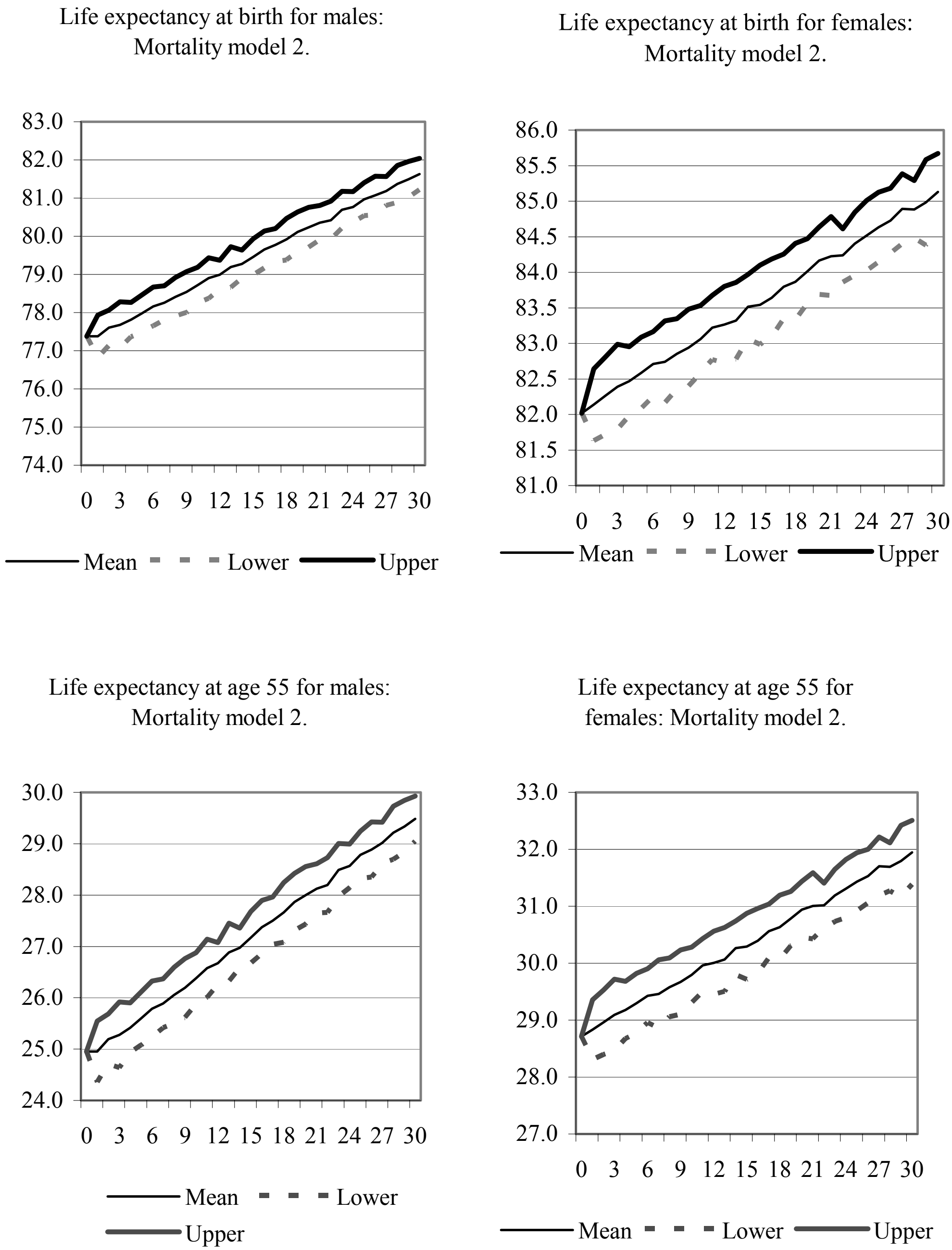
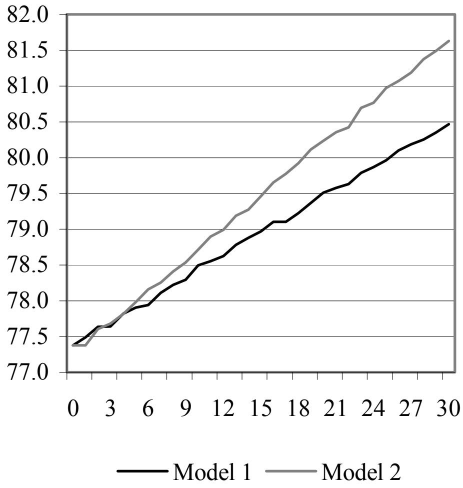
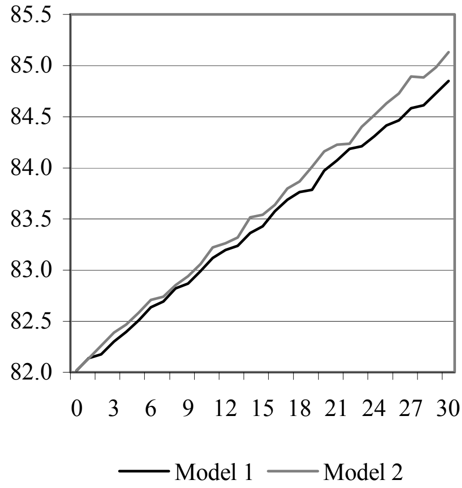


Fig. 3.22 Comparisons of life expectancies for mortality models 1 and 2

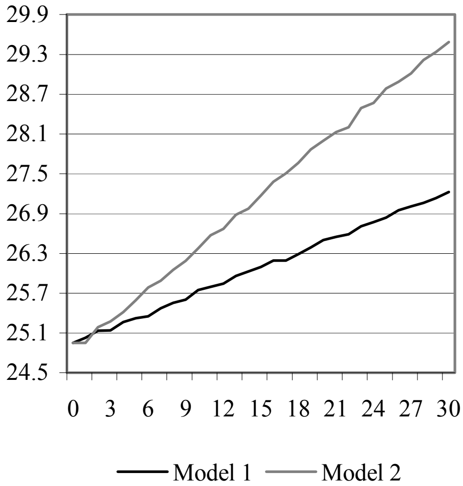
Life expectancies at birth for males, mortality models 1 and 2.



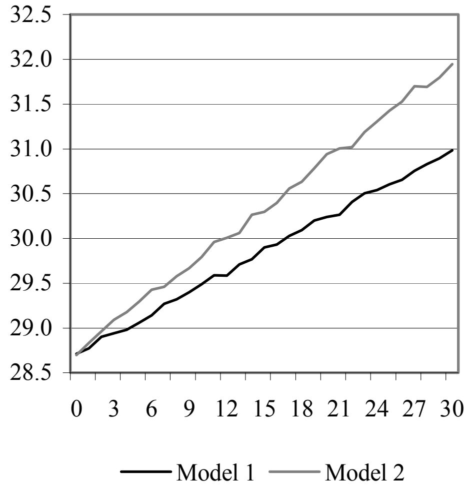
Life expectancies at birth for females, mortality models 1 and 2.



Remaining life expectancy at age 55 for males, mortality models 1 and 2.



Remaining life expectancy at age 55 for females, mortality models 1 and 2.



4 Discussion

The present Statistics Sweden Development Project was undertaken in response to current demographic discussions concerned with estimating the precision of population projections. This is a relatively new area that has brought theory and application of stochastic processes into main focus in demographic contexts (see e.g., Alho, 1990; Goldstein, 2004; Keilman, Pham and Hetland, 2002; Sanderson, Scherbov, O'Neill and Lutz, 2004; Lee, 1998; Long and Hollmann, 2004; Lutz and Goldstein, 2004; Tuljapurkar, 1992). Early contributions are due to Muhsam (1956) and Sykes (1969). The main purpose of stochastic population projections is to enable estimation of prediction intervals for demographic characteristics. An important byproduct of these efforts is that they bring forth an improved understanding of the temporal unfolding of demographic processes (e.g., mortality, marital status, fertility and migration).

Two main approaches distinguish themselves. On the one hand, there is the typical scenario projection where assumptions reflect current demographic, economic and social knowledge (expert panels) and how it is likely to influence the future population with respect to size and structure. On the other, there is the time series approach, which mainly involves modeling future demographic processes on the basis of past and present knowledge of demographic time series. It is fundamental to this approach that the variability observed in the past is projected into the future; an approach taken in the present study. Clearly, in the context of making stochastic population projections, the two modes of thinking are intertwined. It should be noted though that expert panel knowledge usually is of an implicit nature; hence, its implementation is usually of a judgmental nature. In their comprehensive demographic textbook Shryock and Siegel (1976, p. 440) write: *"It is almost commonplace to preface published projections with a statement that it is assumed that the area will not be visited by a war or a natural disaster. Furthermore, no attempt is made ordinarily to allow for future economic fluctuations of a cyclical nature. Of course, the age-sex structure of the population, the marital status and cumulative fertility of its women, and other measures of its current demographic status reflect the impact of such events in the past. The reports presenting the projections do sometimes state that conditions of nearly full-employment are expected to continue or that the gross natural product is expected to increase at a given* *percentage per annum – even though relatively little is known about how variations in these economic phenomena affect population growth."* Bringing together social, economic and demographic knowledge in furtherance of 'better forecasts' still remains a relatively unexplored area.

Throughout the study it has been attempted to apply simplicity and transparency. The study gives focus to standard one-dimensional time series methods that are easily explained and applied. These properties impute transparency. Although upcoming implement-tation necessarily will involve application of more comprehensive models, it is important to maintain simplicity and transparence, as otherwise users may reject the approach on the grounds that it is an incomprehensible black box. The project, it may be added, has benefited from discussions with experts on stochastic population projections and users of projections.

It is important to separate long-term from short-term projections. In the short run, many different models of mortality, fertility and migration usually lead to almost indistinguishable projections. This is true of ordinary scenario projections as well as of stochastic projections. Hence, what may appear as fine-tuned models of mortality, fertility or migration may, in the short run, lead to prediction intervals very similar to those derived from far cruder models. What separates a long-term from a short-term projection, of course, depends on circumstances. Nevertheless, the results of this project suggest that a typical short-term projection extends about five years into the future. But even in the case of long-term projections, different models may lead to numerically similar prediction intervals! This is a finding that lends support to application of *standard* time series models; which is also what can be inferred from studying leading references regarding stochastic population projections.

It is often noted that time series models only are applicable when large time series are available. Because most nations do not boast long demographic time series it might be concluded that stochastic projections only are relevant for industrialized nations (with long demographic time series). This could be a premature conclusion. The need to incorporate demographic variability in projections for developing nations is as important as for industrialized nations. Since available research suggests that for many different countries, mortality, fertility and migration follow similar time series models there is good reason for applying similar methods for nations with sparse demographic data. This, of course, is an important demographic research area that requires further research.

Several papers supporting the present study were written. Three of them have been included in the present report. The two first papers (mortality models 1 and 2) deal with modeling survival so that it becomes a random function. The two models, as well as variations of the two methods, are relatively simple and appear to work well. The third paper deals with an exponential model of fertility. The advantage of this model is that it gives a rather precise representation of low fertility schedules. It has been compared to the standard approach of using the gamma probability density as a scaled model of age-specific fertility. Both the exponential model and the gamma density stand recommended for modeling the age-pattern of fertility in the context of population projections.

References

Alho, J (1990). Stochastic methods in population forecasting. *International Journal of Forecasting*, 6,4, 521-530.

Box, G.E.P. & Jenkins, G.M (1970) Time *Series Analysis, Forecasting and Control,* San Francisco, Holden-Day (revised edn. published 1976).

Brass, W (1971). *On the Scale of Mortality* in Biological Aspects of Demography, ed. By William Brass, pp. 69-110.

Brass et al. (1968). *The Demography of Tropical Africa*. Princeton: Princeton University Press.

Brass, W (1974). Mortality Models and their Uses in Demography. *Transactions of the Faculty of Actuaries*, Vol. 33, No. 2, pp. 123-142.

Brass, W (1975). *Methods for Estimating Fertility and Mortality from Limited and Defective Data.* The Carolina Population Center. University of North Carolina at Chapell Hill.

Carrier, N. H. & Goh, T.J (1972). The Validation of Brass's Model Life Table System. *Population Studies*, Vol. 26, No. 1, pp. 29-51.

Chatfield, C (1999). *The Analysis of Time Series: An Introduction.* London: Chapman & Hall/CRC.

Chiang, C. L (1962). *Introduction to Stochastic Processes in Biostatistics*. New York: Wiley & Son.

Congress of the United States, Congressional Budget Office (2001). *Uncertainty in Social Security's Long-Term Finances: A Stochastic Analysis*. Washington, D.C.

Cox, D.R. & Miller, H.D (1965). *The Theory of Stochastic Processes*. London: Methuen & Co Ltd.

Goldstein, J (2004). Simpler Probabilistic Population Forecasts: Making Scenarios Work. *International Statistical Review*, 72, 93-106.

Gottman, J.M (1981). *Time-series Analysis*. Cambridge: Cambridge University Press.

Hamilton, J.D (1994). *Time Series Analysis*. Princeton: Princeton University Press.

Hartmann, M (1989). Variations in the Age-Pattern of Fertility in Sweden Around 1986. *R&D Report 1989:9.* Statistics Sweden.

Hartmann, M (2006). Migration in Stochastic Population Projections. To appear in *Scandinavian Population Studies*, Vol. 14.

Hill, K. & Trussell, J (1977). Further Developments in Indirect Mortality Estimation. *Population Studies*, Vol. 31, No. 2, pp. 313-334.

Hörmander, Lars (1963). *The Analysis of Linear Partial Differential Operators I: s Distribution Theory and Fourier Analysis.* Berlin: SpringerVerlag.

Keilman, N., Pham, D.Q. & Hetland, A (2002). Why population forecasts should be probabilistic – illustrated by the case of Norway. *Demographic Research*, 6, Article 15. Max Planck Institute for Demographic Research.

Keilman, N. & Dind Quang Pham (2004). Time Series Based Errors and Empirical Errors in Fertility Forecasts in the Nordic Countries. *International Statistical Review*, 72,1,5-18.

Kendall, M (1976). *Time-Series*. London. Charles Griffin and Company Ltd.

Keyfitz, N (1978). *Introduction to the Mathematics of Population*. Addison-Wesley Publishing Company.

Lee, R.D (1998). Probabilistic approaches to population forecasting, in: Lutz, W., Vaupel, J.W. and Ahlburg, D.A. (eds). Frontiers of Population Forecasting, *Population and Development Review*, 24,156-190.

Long, F. & Hollmann, F.W (2004). Developing Official Stochastic Population Forecasts at the US Census Bureau. *International Statistical Review*, 72, 201-208.

Lutz, W. & Goldstein, J (2004). How to Deal with Uncertainty in Population Forecasting? *International Statistical Review*, 72, 1, 1-4.

Lutz, W. & Scherbov, S (2004). Probabilistic Population Projections for India with Explicit Consideration for the Education-Fertility Link. *International Statistical Review*, 72, 1, 81-92.

Muhsam, H.V (1956). The Utilization of Alternative Population Forecasts in Planning. *Bulletin of the Research Council of Israel*, 5(2-3):133-146.

Sandorson, W.C., Scherbov, S., O'Neill, B. & Lutz, W (2004). Conditional Probabilistic Population Forecasting. *International Statistical Review*, 72, 157-166.

Shryock, S., Siegel, J.S. & Associates (1976). *The Methods and Materials of Demography*. Condensed Edition ed. by Stockwell, E. New York: Academic Press Inc.

Statistical Office of the European Communities (1999). *Working Paper No. 41*, 1999. Brussels. European Communities.

Strandell, G (2006). *An Exponential Fertility Model with Three Parameters*. Symposium i anvendt statistik, Informatik og Matematisk Modellering, Danmarks Tekniske Universitet, Danmarks Statistik.

Sykes, Z.M (1969). Some Stochastic Versions of the Matrix Model for Population Dynamics. *Journal of the American Statistical Association*, 64 (325):111-113.

Tuljapurkar, S (1992). Stochastic population forecasts and their uses. *International Journal of Forecasting*, 8, 385-391.

Yaglom, A.M (1962). *An Introduction to the Theory of Stationary Random Functions.* Englewood Cliffs: Prentice-Hall, Inc.

Technical Documentation

1 Introduction

1.1 Stochastic processes

In applications, any variable system subject to random influences constitutes a stochastic process. Briefly, a stochastic process unfolds when a random variable x_t, the argument of which is time, can be observed for a sequence of time-points $t_1, t_2, ..., t_n$. For each t, the distribution function for the random variable x_t is $F_t(x) = P(x_t < x)$. The argument t can also be continuous. For simplicity we assume that t is discrete. Observations $x_1, x_2, ..., x_n$ plotted as a function of time t = 1, … , n, are said to give a realization of the process. A particular value at t = t′, $x_{t'}$, is known as a section of the process. We also refer to $x_1, x_2, ..., x_n$ as a time series or random sequence.

Generally, for a process to be completely specified, it is required that for any sequence of time-points $t_1, t_2, ..., t_n$, and arbitrary n, the finite multidimensional distribution function

$$F_{t_1, t_2, ..., t_n}(x_1, x_2, ..., x_n) = P(x_{t_1} < x_1, x_{t_2} < x_2, ..., x_{t_n} < x_n)$$

is known. In practice, this requirement is rarely met. A simplifying assumption is that the process is stationary. This involves that if for any n, time points $t_1, t_2, ..., t_n$ are translated by the same arbitrary displacement τ, then

$$F_{t_1+\tau, t_2+\tau, ..., t_n+\tau}(x_1, x_2, ..., x_n) =$$
$$F_{t_1, t_2, ..., t_n}(x_1, x_2, ..., x_n)$$

A process that satisfies this requirement is known as strictly stationary (Yaglom, 1962 pp. 9-12). The arbitrary displacement τ is known as the time lag. Letting n = 1, we have, for all t,

$F_t(x) = F(x)$ so that, in effect, $E[x_t] = \mu$ and Vary $[x_t] =$

$E[x_t - \mu]^2 = \sigma^2(t) = \sigma^2$ do not depend on t. For a strictly stationary process, then, the mean and variance do not depend on historical time t. Moreover, the covariance between terms τ time units apart (in any direction) is $\gamma(\tau) = E[(x_t - \mu)(x_{t+\tau} - \mu)]$ and only depends on τ. The function $\gamma(\tau)$ is known as the autocovariance function and for $\tau = 0$ yields the variance $\gamma(0) = \sigma^2$. The corresponding autocorrelation function is $\rho(\tau) = \gamma(\tau)/\gamma(0)$.

A process satisfying that $E[x_t] = \mu$, Var $[x_t] = \sigma^2(t) = \sigma^2$ are independent of t and, moreover, $\gamma(\tau) = E[(x_t - \mu)(x_{t+\tau} - \mu)]$ only depends on lag τ is known as stationary in the weak sense, or as second-order stationary (this is also known as the correlation theory of time series) (Yaglom, 1962). Notice that here we do not require knowledge of the finite dimensional distribution functions, but only that the process has time-invariant first and second-order moments.

A discrete-time process is called a purely random process if it consists of a sequence of random variables z_t that are mutually independent and identically distributed. This process has constant mean and variance, and correlation function $\rho(k) = 1$ if $k = 0$ and $\rho(k) = 0$ otherwise. A purely random process is also known as *white noise*. A white noise term z_t is known as an *innovation*.

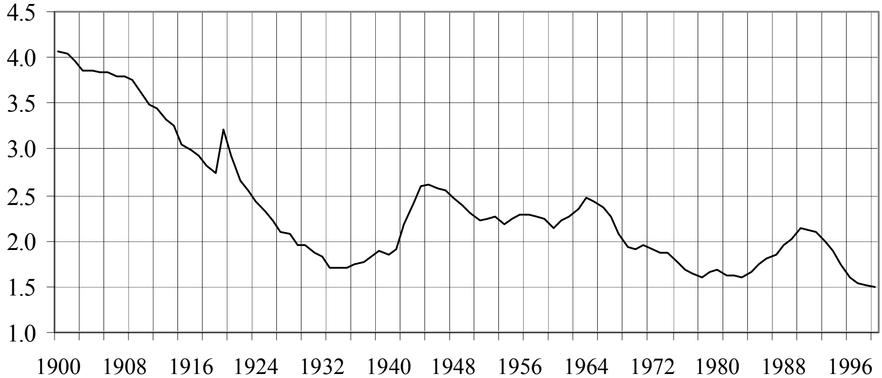
First and second-order ergodicity implies that the autocorrelation function $\rho(\tau) = \gamma(\tau)/\gamma(0)$ rapidly approaches zero as $\tau \to \infty$ (see e.g., Yaglom, 1962 pp. 16-22 for a detailed discussion). For such processes, first and second order moments can be estimated using time-averages from a single realization of the process. This is an important simplification because "social science processes" usually have only one realization. In what follows, stationarity will mean second-order stationarity. In addition, it will be assumed that second-order ergodicity holds for the processes being discussed. This implies that means and variances can be estimated from sufficiently long time-averages obtained from a single realization of the process being studied.

1.2 Demographic processes

Although occasionally manifestations of mortality, fertility and migration are referred to as *processes* in the demographic literature, this does not necessarily imply that authors have in mind *stochastic processes* in the mathematical-statistical sense as discussed in Section 1.1. In fact, in ordinary demographic settings, studies of mortality, fertility and migration rarely involve application of time series models. All the same, a fundamental reason for viewing time sequences of demographic variables as realizations of stochastic processes is that, generally, they display deep-seated stochastic features and hence never can be predicted with total precision, -- especially not many years into the future.

Fig. 1.1 illustrates the time-pattern of the total fertility rate (TFR) for Sweden during the 20th century. Because the time-pattern for each decade is different from any other, it is evident that the curve unfolds with considerable uncertainty. Indeed, while the overall features of the time-pattern can be broadly explained in terms of social, economic, political and historical outlooks (socio-economic scenarios), nevertheless, these explanations are fraught with so much uncertainty and lack of predictive power that they fail in depicting with reasonable precision the future course of TFR. In particular, rational explanations could hardly be found for the many miniscule undulations in the curve. In other words, although some of the features of the time-pattern undoubtedly could be explained in rational terms, others could not, and taken as a whole this means that the curve moves into the future in an unpredictable manner. Nevertheless, if we were only interested in predicting the value of TFR a few years into the future, we could do so with much better precision than if our goal were to estimate it many decades into the future. The main reason for this is the positive autocorrelation in the TFR-series.

Fig. 1.1. Time-pattern for total fertility rate: Sweden 1900-1999



Consider a stationary process for which the autocovariance $\gamma(\tau) = E[(x_t - \mu)(x_{t+\tau} - \mu)]$ is positive for $|\tau| < T$ (it is assumed that $\gamma(\tau) \approx 0$ for $|\tau| > T$, T being an appropriate time stretch). Heuristically, this implies that if the section x_t is above the mean of the process then, likely, neighboring sections $x_{t+\tau}$ are also above the mean for reasonably small τ. Evidently, a time series that is positively autocorrelated has sequences of neighboring sections that hang together in runs either below or above the mean. Stated differently, neighboring sections often change moderately, for which reason x_t often performs as a good first approximation of x_{t+1}. This is the predictive *tour de force* of positively correlated time series. On the other hand, if the series is negatively autocorrelated then this means that there must be frequent jumps in x_t above and below the mean of the process (a zigzag pattern). Whether this induces lesser predictability in the series, of course, depends on the particular process.

Observations that are markedly positively or negatively correlated behave in a manner poignantly different from that of independent observations. Data that spring from the real world are usually correlated. Gottman (1981 p. 41) writes: *"Since it is so difficult to disassociate observed events from some sort of idea of occurrence in time, it seems remarkable that most of the body of statistical methodology is devoted to observations for which the temporal sequence is of no* *importance. Classical statistical analysis requires independence, or at least zero correlation, among observations."*

1.3 How are extrapolation specifications determined?

The (usually high) positive autocorrelations in demographic time series often make short-term extrapolations relatively safe. While there are exceptions to this common rule, migration is one, changes from year to year are often so small that it takes little expertise (statistical, demographic, or otherwise) to make a sound prediction some one or two years into the future. It is probably this commonly noticed feature that entices forecasters to look for ways and means of safely predicting the population over longer future periods.

The d-years lead-time model $\hat{x}_{t+d} = \bar{x}$, where $\bar{x}$ is a time-average of sections of the relevant variable k years back in time, that is,

$$\bar{x} = [1/(k+1)] \sum_{0}^{k} x_{t-j},$$

is a common and intuitive specification in Statistical Bureaus. For small d and k, it often performs well. Nevertheless, in a long-term perspective, such a method is bound to fail. In the first place, even if the process is stationary, a short time-average is not sufficient for reasonably precise estimation of its mean. Certainly, some three, four or five observed values of the process necessarily fall short of taking advantage of ergodicity of the mean. Hence, $\bar{x}$ does not estimate the mean of the observed process but principally the level of x_t in the recent past; it is an intuitive estimate of what is believed to be the level of the relevant variable a few years into the future. Its impressionistic validity does not reflect assumptions concerning stationarity (or non-stationarity). In the second place, because the future course of the observed process is extrapolated as a straight line, which for each year during the period of projection yields one and only one value, namely $\bar{x}$, the temporal variability that is so characteristic of stochastic processes remains unincorporated in the projection process. A third and equally important consideration is that such a method may stand contradicted by the autocorrelations of the data (see e.g., Yaglom, 1962 pp. 111-112).

1.4 Making the cohort projection method stochastic

If mortality, fertility and migration are perceived as stochastic processes then the projection method turns into a mechanism for simulating their conveniently modeled realizations. This is the time series approach to making stochastic population projections. Realizations of mortality, fertility and net-migration become inputs to a system that consists of an initial age and sex-distribution. By repeating the procedure a large number of times (repeated simulations), confidence limits can be estimated for the total population as well as for its characteristics during each year of the projection period.

Prior to making such computations, it is necessary to determine empirically which time series models should be adopted. Once the models have been chosen with empirically estimated parameters, ensuing confidence intervals are conditional in the sense that they are true only if the chosen models are true. There is, it must be emphasized, no such thing as absolute confidence intervals that remain true in all circumstances. It is also important to realize that when stochastic models of mortality, fertility and migration are embedded in an initial age distribution, the resulting simulations may be somewhat insensitive to the choice of models; a fine-tuned model of net-migration could easily give simulations that from a practical point of view are the same as those derived from a much cruder model of net-migration. Hence, it is not until a model has been used in the simulation process that its usefulness can be fully appreciated (in the context of making stochastic population projections).

2 Stochastic Models of Demographic Processes

2.1 Autoregressive models

Although a plethora of time series models can be brought to bear on observed time series of demographic variables, practicality imposes simplicity. Among the models we have studied, standard autoregressive models of the first and second order provided the most convenient description of migration and fertility. These models are briefly outlined below.

Let Z_t be a purely random process with E[Z_t] = 0 and variance Var[Z_t] = σ_z^2. Then {X_t} is said to be an autoregressive process of order p if

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \ldots + \alpha_p X_{t-p} + Z_t$$

This process is abbreviated AR(p). Here X_t is a centered variable, that is, for all t, E[X_t] = 0. The model involves that X_t is generated by means of linear regression on sections at times t-1, t-2, … , t-p as well as a random chock (innovation) Z_t at time t. A numerically convenient specification is

$$X_t - \mu = \qquad (2.1)$$

$$\alpha_1(X_{t-1} - \mu) + \alpha_2(X_{t-2} - \mu) + \ldots + \alpha_p(X_{t-p} - \mu) + Z_t$$

where E[X_t] = μ. For (2.1) to define a stationary process, the parameters α_i must satisfy certain regularity conditions. In practical applications, AR(1) and AR(2) are the most frequently used autoregressive models.

The AR(1) model is

$$X_t = \alpha X_{t-1} + (1-\alpha)\mu + Z_t \tag{2.2}$$

which is stationary if $|\alpha| < 1$ with mean $E[X_t] = \mu$ and variance

$$\mathrm{Var}[X_t] = \sigma_x^2 = \frac{\sigma_z^2}{1-\alpha^2}. \tag{2.3}$$

where $\mathrm{Var}[Z_t] = \sigma_z^2$. The process has autocovariance function

$$\gamma(k) = \alpha^{|k|}\sigma_x^2 \tag{2.4}$$

at lag k and autocorrelation function

$$\rho(k) = \alpha^{|k|} \tag{2.5}$$

for lags $k = 0, \pm 1, \pm 2, \ldots$.

The AR(2) model is

$$x_t = \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + (1-\alpha_1 - \alpha_2)\mu + e_t \tag{2.6}$$

For the AR(2) model to be stationary it is necessary that

$$\alpha_1 + \alpha_2 < 1 \tag{2.7}$$

$$\alpha_1 - \alpha_2 > -1,$$

and

$$\alpha_2 > -1$$

The corresponding auxiliary equation is

$$y^2 - \alpha_1 y + \alpha_2 = 0$$

the roots of which are real if $\alpha_1^2 + 4\,\alpha_2 > 0$, otherwise complex. When the roots are real the autocorrelation function decreases exponentially with lag k. When the roots are complex the autocorrelation function is a damped harmonic

$$\rho(k) = \left(\sqrt{(-\alpha_2)}\right)^k \frac{\cos(k\beta + \Psi)}{\cos\Psi} \tag{2.8}$$

where Ψ is a constant chosen to agree with

$$\rho(1) = \frac{\alpha_1}{1 - \alpha_2}, \tag{2.9}$$

the autocorrelation at lag 1, and

$$\tan\beta = \frac{\alpha_1}{\sqrt{(-(\alpha_1^2 + 4\alpha_2))}}$$

The variance of the AR(2) process is

$$\sigma_x^2 = \frac{1 - \alpha_2}{(1 + \alpha_2)((1 - \alpha_2)^2 - \alpha_1^2)}\sigma_e^2 \tag{2.10}$$

where $\mathrm{Var}[e_t] = \sigma_e^2$ (Cox and Miller, 1965 pp. 282-283).

2.2 A numerical illustration using AR(1) as a model of TFR

The total fertility rate for Sweden in 2000 was TFR = 1.56. Fig. 2.1 shows 10 realizations of TFR, based on AR(1) with start x_0 = TFR = 1.56 in 2000, taking the process 30 years into the future with the same variance as observed during 1950-2000 (the reference period for the extrapolation). The estimated standard deviation for the observed series (1950-2000) is $\hat{\sigma}_x = 0.28$ and estimated mean is $\hat{\mu}$ = 1.94. In this example it is assumed that the future mean value of the TFR process is μ = 1.90. The formula (or forecasting equation as it is also called) extrapolating the process m years into the future is $\tilde{x}_{t+m} = \alpha^m (x_t - \mu) + \mu$ which for $|\alpha| < 1$ converges toward μ. For high values of α the convergence is slow. This is demonstrated below (fig. 2.2). The estimated forecasting equation ($\hat{\alpha}$ = 0.972) is $\tilde{x}_{t+m} = 0.972^m(-0.34) + 1.90$. The estimated standard deviation of white noise is $\hat{\sigma}_e$ = 0.08.

As noted, convergence toward μ provided by the forecasting equation[6] $\tilde{x}_{t+m} = \alpha^m(x_t - \mu) + \mu$ is slow if $\alpha \approx 1, |\alpha| < 1$. With the present estimate $\hat{\alpha}$ = 0.972, it would take about one hundred years for $\tilde{x}_{t+m}$ to reach μ = 1.90. This illustrates, in the light of mathematical modeling, why short-term is safer than long-term forecasting. It should also be noted that the extrapolation trace is not a straight line, but a curve $\alpha^m(x_t - \mu) + \mu$ where the term $x_t - \mu$ is exponentially suppressed, with limiting value μ as $m \to \infty$.

[6] Forecasting equations of this nature provide dynamic forecasts as illustrated in fig. 2.2. While such forecasts may be a considerable aid in determining realistic future scenarios for population forecasts, computations enabling estimation of prediction intervals rely on pseudo-realizations as shown in fig. 2.1.

Fig. 2.1 Observed TFR for 1950-2004 and 10 AR(1) realizations, 2000-2030

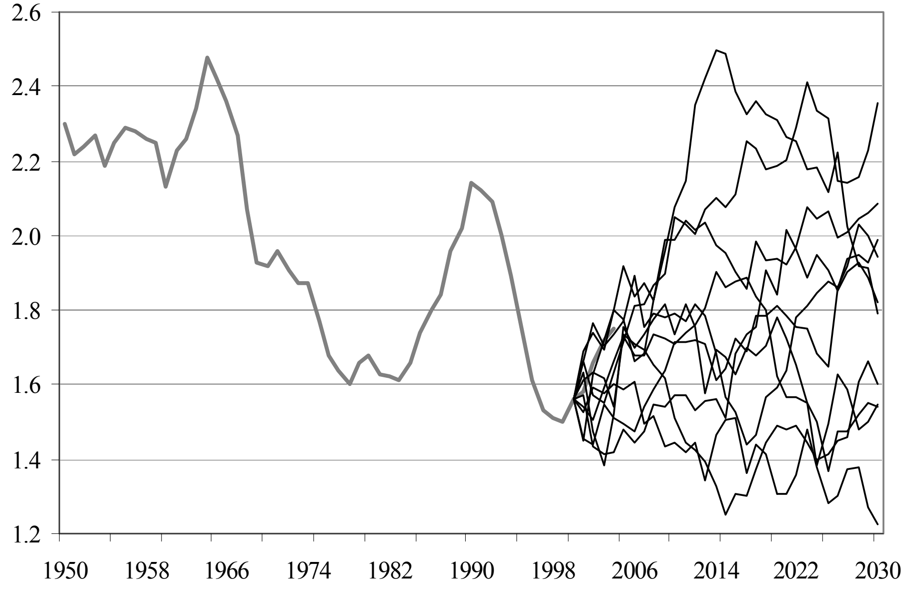
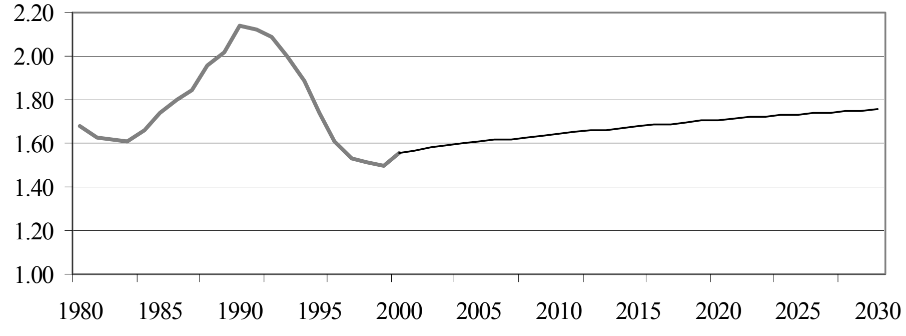


Fig. 2.2. Observed TFR for 1980-2000 and AR(1) forecasting trace 2001-2030



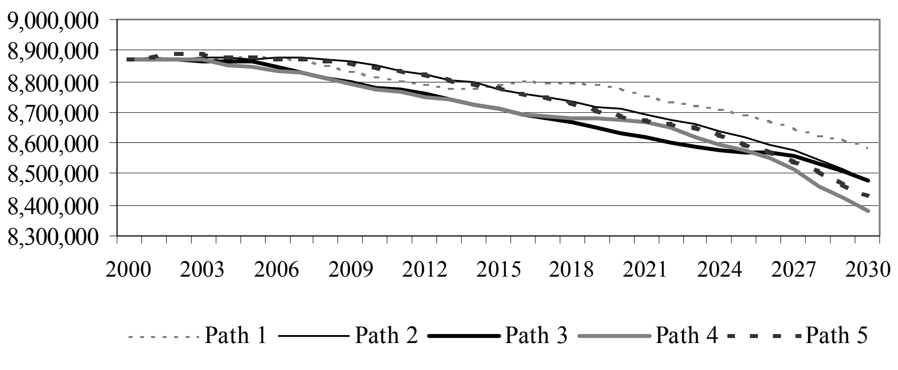
2.3 Numerical example illustrating TFR as random walk with reflecting barriers

Fig. 2.3 shows how the Swedish 2000 both sexes population changes over a thirty-year period when mortality is that of the 2000 life table, zero net-migration and TFR varies as a random walk with reflecting barriers. The paths in fig. 2.3 show the result of five individual runs (realizations) of the random walk.

The random walk is such that TFR in calendar year t is $x_t = \mathrm{TFR}(t)$ with $x_t = x_{t-1} + e_t$. The starting value for the process is $x_{2000} = 1.56$ (TFR in the year 2000). If x_t reaches a ceiling of 2.0 it becomes $x_t = 2.0 + e_t$ and if it x_t reaches a floor of 1.5 it becomes $x_t = 1.5 + e_t$. The error terms are independent and normally distributed with zero mean and standard deviation σ_e = 0.10 (empirically determined from the unfolding of TFR during the 20th century).

While the example is hypothetical, it illustrates how to gain insight into the importance of TFR as a factor behind population growth, when mortality and net-migration are held constant. The variability in the projected population, then, solely reflects the impact of fertility. In the year 2030, the population size varies between a minimum of 8.38 and a maximum of 8.59 million; a range of 207,073 persons.

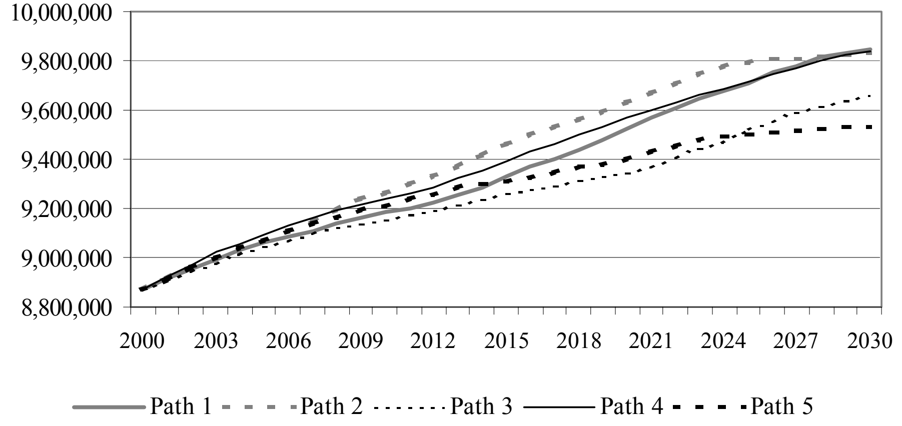
Fig. 2.3. Five repetitions of population growth during a thirty-year period when TFR varies as a random walk with reflecting barriers



As matter of computational detail, it should be noted that in the above example a fixed set of appropriately chosen normalized age-specific fertility rates is scaled so as to yield the given TFR in calendar year t. In reality, of course, the age-pattern of fertility is likely to change during the period of projection.

The example is also instructive of the population projection technique being an exploratory or insight-providing tool. The diagnostic value of the technique is further illustrated in fig. 2.4 where it is assumed that net-migration each year is 20,000 persons.

Fig. 2.4. Five repetitions of population growth during a thirty- year period when TFR varies as a random walk with reflecting barriers, and when it is assumed that net-migration is 20,000 persons per year



The population in 2030 now varies between 9.84 and 9.53 million, a range of 311,624. In addition, it should be noted that net-migrants are distributed evenly between men and women, and that age-specific net-migration is distributed by means of an estimated (time-invariant) age-distribution around the year 2000. In reality, age-distributions for international in and out-migration change from year to year. Also, the balance between men and women, too, is subject to temporal change. It will be realized therefore that the stochastic approach to making population projections does not dispense with the need to make assumptions concerning future levels, and sex and age-patterns, of demographic processes.

3 Choice of Models

3.1 Net-migration

It is net-migration, that is, the balance between in and out-migration that is added (or subtracted) to the projected population each year during the projection period. This is the *modus operandi* of the projection process when it involves migration. Because net-migration is a volatile process that may change dramatically from period to period, it is not amenable to safe prediction. In fact, typing the net-migration process is difficult simply because of its highly non-stationary character. In the case of Sweden however it has behaved very much like a first order autoregressive process, AR(1), after the 1950s (Hartmann, 2006). Generally, in the case of stochastic population projections net-migration is often implemented as an AR(1) process, even though it is much more complex than that. In this project net-migration is modeled as AR(1) and distributed evenly between men and women, and in accordance with age-specific distributions. These age and sex-distributions are those that prevailed around 2000 in Sweden.

3.2 Mortality

As already noted, in ordinary life table notation, the projection probability at age x is $\pi_x = L(x+1)/L(x)$. This is the expected proportion of those aged x who attain age x+1. If the midyear population aged x is P(x), the expected number of survivors aged x+1 is

$$P(x+1) = \pi_x P(x) \qquad (3.1)$$

The projection probability at age x at time t is $\pi_x(t)$. If P(t; x) is the midyear population aged x in calendar year t, then (3.1) becomes

$$P(t+1; x+1) = \pi_x(t)\, P(t\,; x) \qquad (3.2)$$

Because $\pi_x(t)$ is a function of life table exposures L(t; x), its derivation is the survival function s(t; x) during calendar-year t. Ordinarily it is assumed that

$$L(t; x) = [s(t; x)+s(t; x+1)]/2 \tag{3.3}$$

This is the standard approximation used in this project. Because the projection probabilities derive from the survival function, it was deemed appropriate in the present project to model the survival function as a random function using the Brass logit life table model (see Technical Papers).

3.3 Fertility

In the context of population projections, it is the total fertility rate, TFR, that references fertility. Specifically, the fertility schedule at time t is $\{f_x^t\}$ with

$$f_x^t = TFR_t \sum_{x=\alpha}^{\beta} n_x \tag{3.4}$$

where $\{n_x\}$ is a normalized fertility schedule, that is, $\sum_{x=\alpha}^{\beta} n_x = 1$ and $[\alpha, \beta]$ is the reproductive age-interval for women. The mean value[7] and the variance of $\{f_x^t\}$, by definition, are

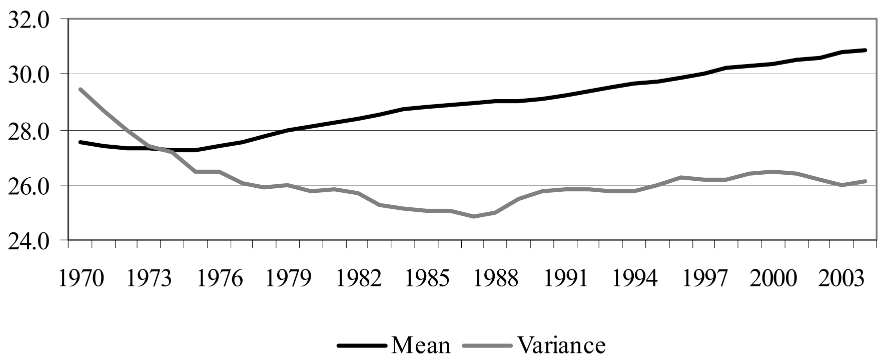
$\mu \approx \sum (x+0.5)\, n_x$ and $\sigma^2 \approx \sum (x+0.5-\mu)^2\, n_x$, respectively. The traditional approach to fertility specifications (in projections) involves letting $\{n_x\}$ be time-invariant and merely scale it appropriately across the projection period. Nevertheless, in reality, the mean age and the variance of the observed fertility schedule $\{f_x^t\}$ change over time. Fig. 3.5 shows mean age and variance for Swedish observed fertility schedules 1970-2004. During this period the mean age of the fertility schedule increased from 27.5 to 30.9 years. From the mid-1970s the variance has remained close to $\sigma^2 \approx 26$ (a highly condensed schedule the variance of which is about half that of historical schedules).

The mean and the variance are two important descriptive indexes of the age-pattern of fertility and for this reason ought to play important roles when specifying age-specific fertility in forecasts

[7] Here x is age last birthday so that exact age is approximately x+0.5.

(clearly one would have to distinguish between long and short-term forecasts). Because of the weak (empirical) correlations between TFR and μ and σ^2, it necessarily involves speculation to specify their future temporal relationships. This, undoubtedly, is one of the major reasons for specifying future fertility in terms of TFR and a normalized fertility schedule with time invariant mean and variance. These circumstances withal, it is important to give some thought to the future age-pattern of fertility. Clearly, it is not very likely, for example, that the mean age of Swedish fertility schedules will remain at about 30 years for however long into the future; historically age-patterns of fertility have changed considerably over time, -- and are likely to do so in the future. This is an issue beyond the scope of the present report. Strandell (2005, see Technical Papers) has demonstrated the magnitude of error in the yearly number of births in Sweden that comes about as a result of applying a fixed fertility schedule during a projection period.

Fig. 3.5. Mean value and variance for observed fertility schedules: Sweden 1970-2004



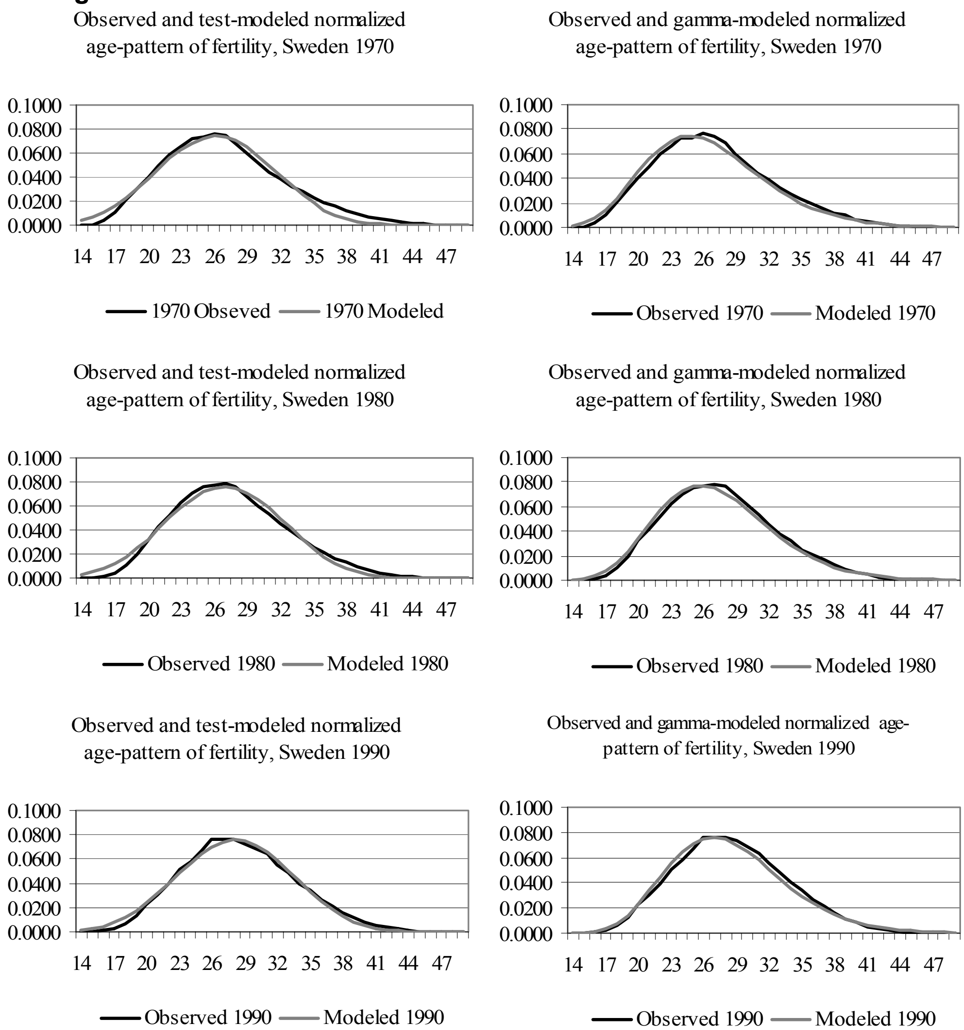
Strandell (2005, see Technical Papers) has suggested the parametric test-function[8]

$$t(x;a,k,m,g) = \frac{a}{\exp\{\frac{k}{1-\frac{(x-m)^2}{g}}\}}, \tag{3.5}$$

[8] A fundamental function in the theory of mathematical distributions (Hörmander, 1963).

$-1 < \frac{x-m}{g} < 1$ as a model of age-specific fertility. Application of (3.5) to the time series of fertility schedules for 1970-2004 shows that parameters g and k may be fixed at g = 32 and k = 17. This modifies (3.5) so that

Fig. 3.5. Normalized age-patterns of fertility modeled by test and gamma functions



$$t(x;a,m) = \frac{a}{\exp\{\dfrac{17}{1-\dfrac{(x-m)^2}{32}}\}} \quad (3.6)$$

becomes a two-parameter model of age-specific fertility. Because (3.5) and (3.6) do not reference a probability distribution, parameter a does not measure the level of fertility in the sense of TFR. Parameter m however is closely related to the mean age of the fertility schedule.

Historically the gamma probability distribution has been widely used to model fertility schedules (Keyfitz, 1968). Here age-specific fertility is modeled

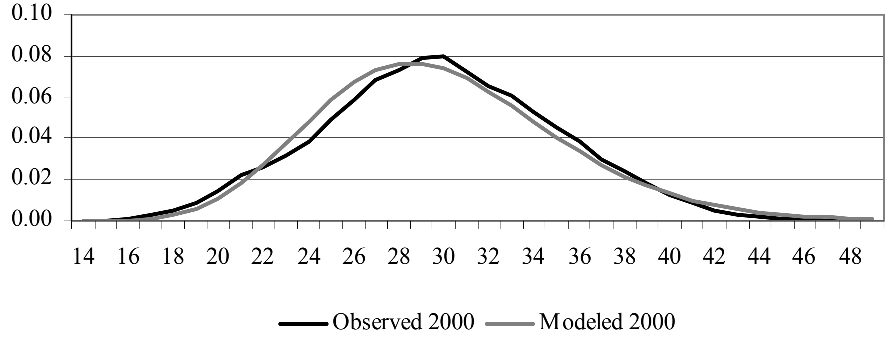
$$g(x;c,k,d) = \mathrm{TFR}\,\frac{c^k}{\Gamma(k)}(x-d)^{k-1}\,e^{-c(x-d)} \quad (3.7)$$

$x > d$. In (3.7) the mean age of the modeled fertility schedule is $\mu = k/c + d$ and the variance is $\sigma^2 = k^2/c$. In (3.7), $\Gamma(k) \approx \sqrt{\frac{2\pi}{k}}\,k^k\,e^{(-k+\frac{1}{12k})}$. It has been shown (Hartmann, 1989) that (3.7) gives a highly accurate fit to Swedish low-fertility age-patterns.

To illustrate the goodness of fit provided by (3.6) and (3.7), fig. 3.5 shows observed normalized fertility schedules and the corresponding modeled ones for 1970, 1980 and 1990. An additional illustration of the test-function (3.5) with g = 32 and with free parameters a, k and m is given in fig. 3.6 which shows observed and modeled normalized fertility for Sweden 2000. Estimated parameters for fig. 3.6 are

$\hat{m} = 30.25$, $\hat{k} = 17.09$ and $\hat{a}/100{,}000 = 19.85$.

Fig. 3.6. Observed and test-function normalized age-pattern of fertility, Sweden 2000



It can be shown that the magnitude of differences between observed and fitted fertility curves in e.g., figs. 3.5-3.6 have no more than negligible effects on projected births. Fig. 3.7 shows estimates of c in (3.7) for calendar years 1970-2004, fig. 3.8 corresponding estimates of k. Regression representations for k and c with respect to the mean value of the fertility schedule μ are

$$k = 1.38\ \mu - 27.58,\ \hat{\sigma}_e(k) = 0.4,\ R^2 = 0.96,$$

$$c = 0.039\ \mu - 0.48,\ \hat{\sigma}_e(c) = 0.02,\ R^2 = 0.87$$

One might be able to take advantage of such regression representations for simulating normalized schedules. However this leaves aside the problem of how to combine the total fertility rate with modeled age-patterns of fertility.

Fig. 3.7. Estimates of c in gamma probability density function for normalized fertility schedules, Sweden, 1970-2004

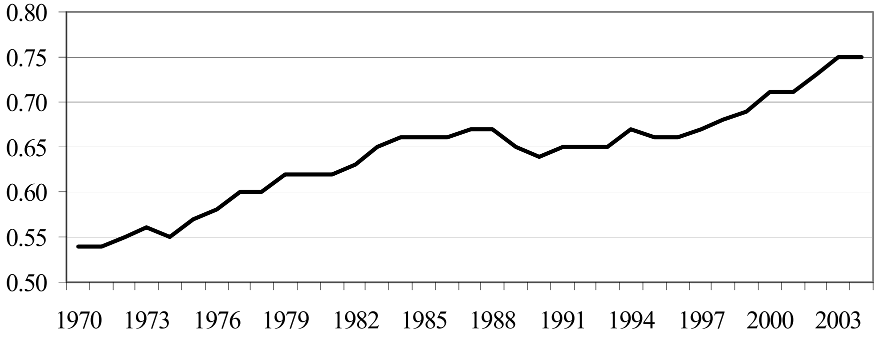
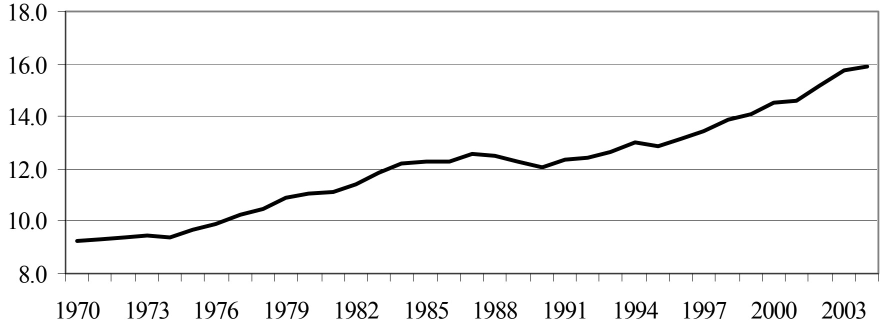


Fig. 3.8. Estimates of k in gamma probability density function for normalized fertility schedules, Sweden, 1970-2004



Mortality model 1:

Mortality in Stochastic Population Projections[9]
Michael Hartmann
Statistics Sweden
March 2005

Abstract

This paper discusses a one-parameter model of the survival function intended for use with stochastic population projections. Section 1 introduces the model, a one-parameter version of the Brass logit life table system, its estimation, and illustrates its precision using Danish life tables. Section 2 references Swedish life tables and illustrates applicability of the model in the perspective of population projections. Section 3 illustrates the model as a means of simulating stochastic survivorship. Section 4 illustrates how the method embeds stochastic survivorship in projected populations. It is concluded that this could be a suitable survival model for use with stochastic population projections.

[9] Paper presented at the Joint Eurostat-UNECE Work Session on Demographic Projections, Vienna, 21-23 September, 2005.

1 Introduction

1.1 Relational survival

Survival functions are linearly related in the sense that for a chosen standard survival function l_x^s, radix one, another survival function l_x can be expressed as

$$\text{logit}\, l_x \approx \alpha + \beta\, \text{logit}\, l_x^s \tag{1.1}$$

where α and β are parameters and $\text{logit}\, p = \ln \frac{1-p}{p}$ with $0 < p < 1$ (see e.g., Brass, 1971, 1974 and 1975; Brass *et al*, 1968; and Hill and Trussell, 1977). It follows from (1.1) that

$$l_x \approx 1/(1 + e^{\alpha}((1 - l_x^s)/l_x^s)^{\beta}) \tag{1.2}$$

To fit the right-hand side of (1.1) to an observed or *a priori* given survival function l_x, $\sum_{x=1}^{\omega} \left[\text{logit}\, l_x - \alpha - \beta\, \text{logit}\, l_x^s \right]^2$ is minimized with respect to α and β. Denoting estimated parameters by $\hat{\alpha}$ and $\hat{\beta}$, fitted survival is

$$\hat{l}_x(\hat{\alpha}, \hat{\beta}) = 1/(1 + e^{\hat{\alpha}}((1 - l_x^s)/l_x^s)^{\hat{\beta}}) \tag{1.3}$$

This unweighted least-squares approach to estimating the parameters in (1.1) is usually adequate for practical work (for alternative estimation procedures, see e.g., Carrier and Goh, 1972). The significance of the parameters is that, relative to the chosen standard survival function, α is related to the level of mortality while β controls the relationship between childhood and adult mortality (see e.g., Brass, 1974 and Brass *et al*, 1968). For (1.3) to provide an adequate approximation to an observed survival function l_x, l_x^s should have approximately the same age-pattern and level of mortality as l_x. In such cases, $\hat{\beta} \approx 1$ (see e.g., Brass, 1971). In applications it is common to chose a standard life table for which it is reasonable to let $\beta = 1$.

In this paper, we adopt the one-parameter model ($\beta = 1$)

$$l_x(\alpha) = 1/(1 + e^{\alpha}(1 - l_x^s)/l_x^s) \tag{1.4}$$

To estimate α, the unweighted sum of squares

$$\sum_{x=1}^{104} \left[l_x - 1/(1 + e^{\alpha}(1 - l_x^s)/l_x^s) \right]^2 \tag{1.5}$$

at ages between 1 and 104 is minimized with respect to α. The main purpose of the paper is to discuss the possibility of letting (1.4) serve as a survival model when making stochastic population projections. Before we discuss this possibility, we turn to an illustration of (1.4) using Danish life tables.

1.2 Illustrative application to Danish life tables

Table 1.1 shows estimates of α when using 1921 Danish male and female survival as standards when fitting (1.4) to Danish male and female survival for the period 1922-51. Fig. 1.1 shows the result of fitting (1.4) to 1936 Danish male survival (using 1921 Danish male survival as a standard). In this experiment, $\hat{\alpha}$ = -0.097 (table 1.1). The life expectancy of the standard is e_o^s = 61.0 years, for observed survival e_o = 62.5, and for the fitted (the estimated model) e_o^f = 62.4 years. This suggests that for a displacement in life expectancy of about one and a half years relative to the standard, (1.4) provides a close fit.

Table 1.1. Estimates of α for 1922-51 using 1921 Danish male and female survival as standards

Year	α Males	α Females	Year	α Males	α Females
1921	0.000	0.000	1937	-0.118	-0.191
1922	0.075	0.083	1938	-0.196	-0.254
1923	0.042	0.024	1939	-0.248	-0.325
1924	0.042	0.027	1940	-0.285	-0.335
1925	-0.020	-0.018	1941	-0.267	-0.335
1926	0.008	-0.008	1942	-0.357	-0.416
1927	0.025	0.042	1943	-0.378	-0.435
1928	-0.003	-0.030	1944	-0.304	-0.354
1929	-0.016	-0.021	1945	-0.274	-0.339
1930	-0.028	-0.055	1946	-0.368	-0.405
1931	0.006	-0.028	1947	-0.444	-0.525
1932	-0.067	-0.068	1948	-0.548	-0.695
1933	-0.117	-0.130	1949	-0.550	-0.675
1934	-0.136	-0.187	1950	-0.564	-0.675
1935	-0.069	-0.098	1951	-0.609	-0.729
1936	-0.097	-0.144			

Fig. 1.2 shows similar graphs for Danish females. Here 1921 Danish female survival serves as a standard. For this standard $e_o^s = 62.5$, for observed 1936 female survival $e_o = 64.6$, and for the fitted $e_o^f =$ 64.4 years.

Fig. 1.1. Observed and fitted survival for Danish males 1936 using 1921 Danish male survival as a standard

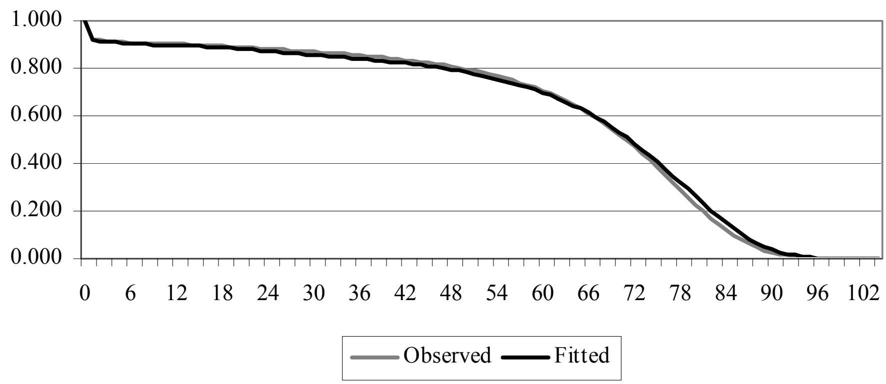
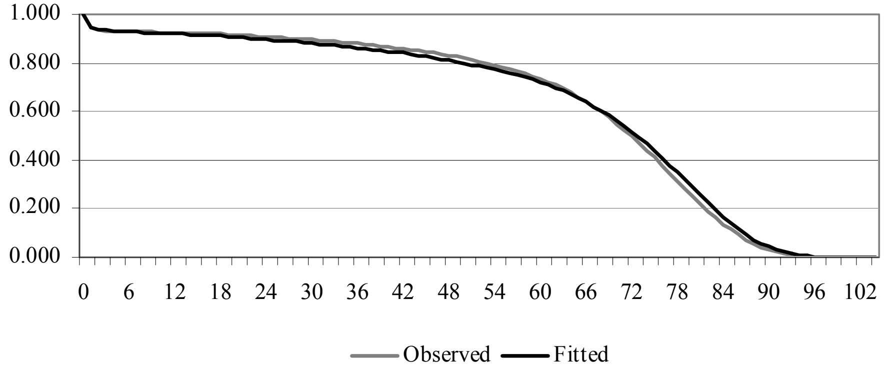


Fig. 1.2. Observed and fitted survival for Danish females 1936 using 1921 Danish female survival as a standard



1.3 Modeling population projection probabilities

The population projection probability, that is, the survival probability for a life aged x to survive to age x+1 is,

$$\pi(x) = L(x+1)/L(x) \tag{1.6}$$

where $L(x)$ are the expected person-years (exposure) in the chosen life table. Because the survival function

$$s(x) = (1-q_0)...(1-q_{x-1})$$

is the product of terms (1-q_i) where q_i, i = 0, … , x-1, is the probability for a life aged i to die before reaching age i+1 (the life table mortality rate), different sets of life table mortality rates may, numerically speaking, lead to virtually indistinguishable survival functions. Since, moreover, $L(x)$ is usually approximated by

L(x) = [(s(x)+s(x+1)]/2,

projection probabilities that only differ marginally may result from rather different underlying mortality rates (as long as they share approximately the same level of mortality). For example, increasing q_x = 0.1 by 10 percent only changes $p_x = 1 - q_x$ from 0.90 to 0.89; a 1.1 percent decline. Indeed, raising all mortality rates in the Swedish 2000 male life table by 10 percent only lowers its life expectancy at birth from 77.4 to 76.4 years; a 1.3 percent change. Stated differently, what in its own right may appear to be a significant change in mortality rates may involve a mere trifling change in survivorship. To illustrate this in more detail, fig. 1.3 shows projection probabilities as determined from the observed life table for Danish males in 1936 as well as those modeled by means of (1.4) using 1921 Danish male survival as a standard. Fig. 1.4 shows the corresponding graphs for Danish females in 1936.

Fig. 1.3. Observed and fitted projection probabilities for Danish males in 1936 (using 1921 Danish male survival as a standard)

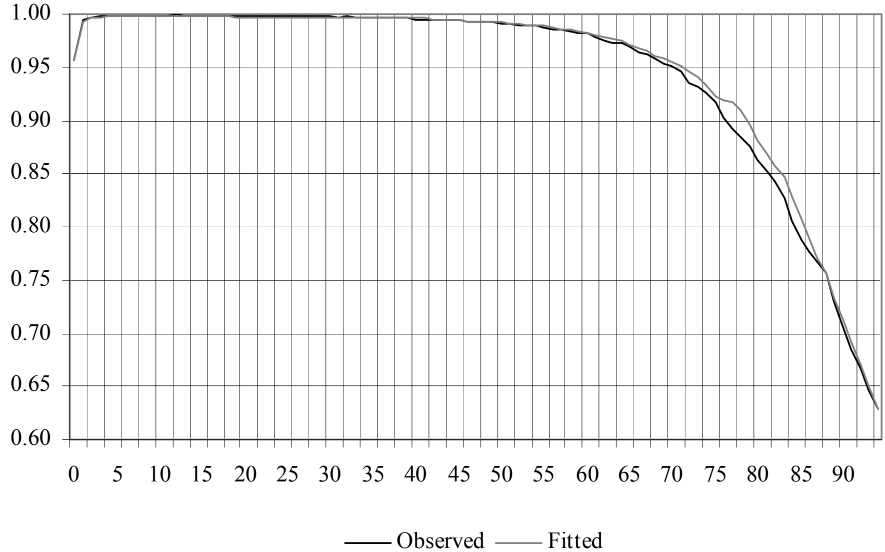
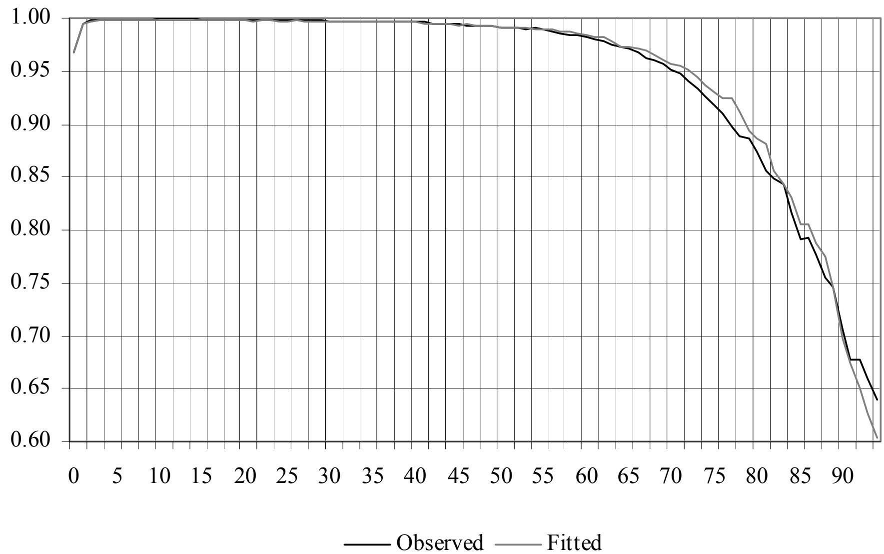


Fig. 1.4. Observed and fitted projection probabilities for Danish females in 1936 (using 1921 Danish female survival as a standard)



2 Projections of the Swedish Population

2.1 Projecting the 1980 population to the year 2000

Table 2.1 shows estimates of α for Swedish males and females 1981-2002 using 1980 Swedish male and female survival as standards. Fig. 2.1 shows the observed and estimated survival functions for Swedish males in the year 2000. The life expectancy for observed survival is $e_o = 77.4$ and for fitted $e_o^f = 77.0$ years. The life expectancy for the standard is $e_o^s = 72.8$ years. Although here the difference in life expectancy between the standard and the modeled survival functions is about 4.5 years, the fit remains close. It is in place to note that when the difference in life expectancy between the standard and the survival to be modeled is small (a difference of some one or two years), the two curves can hardly be distinguished in a graph. Fig. 2.2 shows projection probabilities for observed and modeled survival functions for Swedish males in the year 2000. Once again, it is apparent that the differences between the observed and modeled projection probabilities are marginal. Notwithstanding the smallness of the differences, their computational significance can only be gauged *objectively* if the Swedish population is projected using partly observed, partly modeled probabilities.

Fig. 2.1. Observedd and estimated survival functions for Swedish males in the year 2000 using 1980 Swedish male survival as a standard

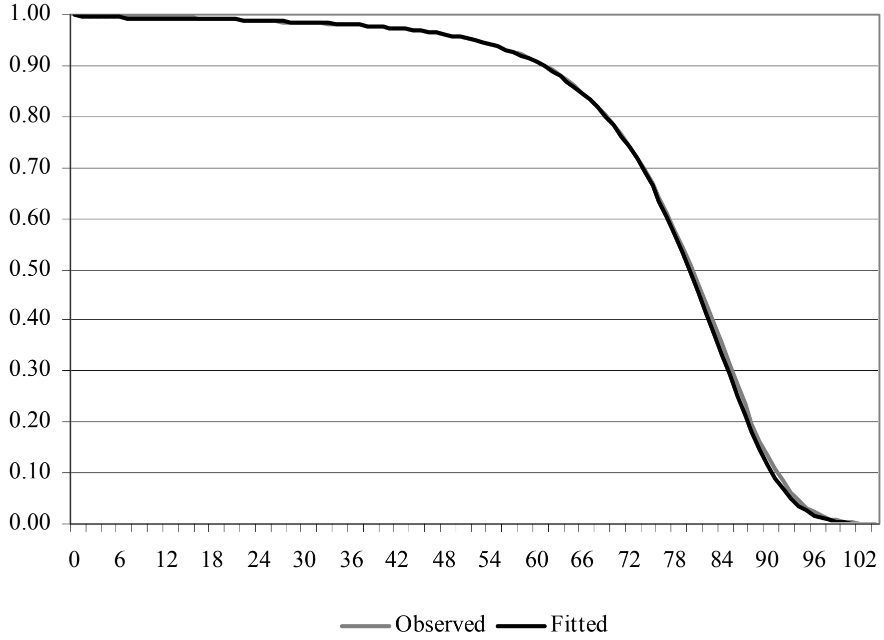


Table 2.1. Estimates of α for Swedish males and females 1981-2002 using 1980 Swedish male and female survival as standards

Year	Males	Females	Year	Males	Females
1980	0.000	0.000	1992	-0.325	-0.282
1981	-0.029	-0.036	1993	-0.345	-0.281
1982	-0.069	-0.078	1994	-0.419	-0.374
1983	-0.100	-0.122	1995	-0.431	-0.375
1984	-0.128	-0.153	1996	-0.475	-0.389
1985	-0.119	-0.129	1997	-0.500	-0.432
1986	-0.147	-0.171	1998	-0.531	-0.451
1987	-0.174	-0.191	1999	-0.560	-0.441
1988	-0.173	-0.162	2000	-0.560	-0.464
1989	-0.259	-0.259	2001	-0.639	-0.474
1990	-0.261	-0.233	2002	-0.660	-0.478
1991	-0.276	-0.258			

Fig. 2.2. Observed and fitted projection probabilities for Swedish males in 2000 using 1980 Swedish male survival as standard

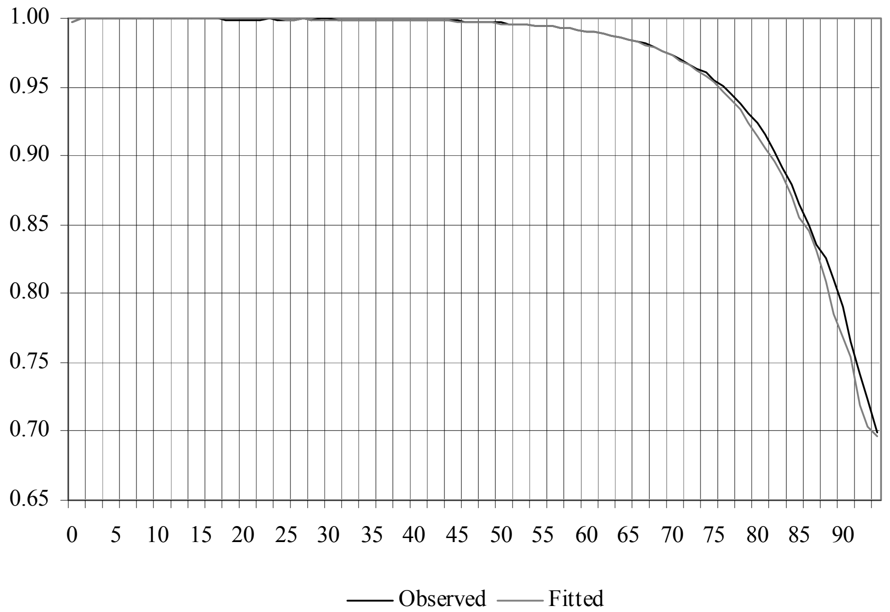


Table 2.2 shows the results of projecting the Swedish 1980 midyear population 30 years into the future using partly projection probabilities calculated directly from the 2000 survival functions for males and females, partly by using projection probabilities derived from the modeled 2000 survival functions for males and females. For example, modeled 2000 male survival is

$$\hat{l}_x = 1/(1 + e^{-0.560}(1 - l_x^s)/l_x^s)$$

where l_x^s denotes 1980 Swedish male survival (see table 2.1).

The population is projected with the assumption that the total fertility rate is constant at TFR = 2.0, zero net-migration and projection probabilities corresponding to the observed 2000 life table (Projection 1) and the modeled 2000 life table (Projection 2). Projection 1 and Projection 2 confirm that differences between observed and modeled projection probabilities are of no consequence for the projected population sizes. The differences in percent between the two projections are shown in table 2.2. The two sets of projection probabilities also lead to similar age distributions. For example, in Projection 1 in the year 2010, the proportions aged 65+ are 17.3 and

20.8 percent for males and females and 17.0 and 20.8, respectively, in Projection 2. In the light of a population projection or forecast, these are insignificant differences.

Table 2.2. Projections of the Swedish 1980 midyear population 30 years into the future assuming a total fertility rate of 2.0, zero net-migration and projection probabilities corresponding to the observed and the modeled 2000 life table

Year	Projection 1			Projection 2			Percent error		
	Both sexes	Males	Females	Both sexes	Males	Females	Both sexes	Males	Females
0	8,310,335	4,117,603	4,192,732	8,310,335	4,117,603	4,192,732	0.00	0.00	0.00
1	8,361,945	4,143,226	4,218,719	8,359,793	4,141,653	4,218,140	0.03	0.04	0.01
2	8,410,154	4,167,237	4,242,917	8,405,992	4,164,182	4,241,810	0.05	0.07	0.03
3	8,454,838	4,189,584	4,265,254	8,448,829	4,185,149	4,263,680	0.07	0.11	0.04
4	8,495,973	4,210,247	4,285,726	8,488,266	4,204,519	4,283,747	0.09	0.14	0.05
5	8,533,625	4,229,259	4,304,366	8,524,369	4,222,329	4,302,040	0.11	0.16	0.05
6	8,567,901	4,246,674	4,321,227	8,557,242	4,238,631	4,318,611	0.12	0.19	0.06
7	8,599,005	4,262,596	4,336,409	8,587,082	4,253,523	4,333,559	0.14	0.21	0.07
8	8,627,170	4,277,137	4,350,033	8,614,108	4,267,105	4,347,003	0.15	0.23	0.07
9	8,652,651	4,290,423	4,362,228	8,638,575	4,279,497	4,359,078	0.16	0.25	0.07
10	8,675,708	4,302,582	4,373,126	8,660,730	4,290,836	4,369,894	0.17	0.27	0.07
11	8,696,621	4,313,753	4,382,868	8,680,862	4,301,259	4,379,603	0.18	0.29	0.07
12	8,715,634	4,324,054	4,391,580	8,699,209	4,310,876	4,388,333	0.19	0.30	0.07
13	8,732,981	4,333,601	4,399,380	8,716,004	4,319,806	4,396,198	0.19	0.32	0.07
14	8,748,867	4,342,494	4,406,373	8,731,439	4,328,142	4,403,297	0.20	0.33	0.07
15	8,763,396	4,350,778	4,412,618	8,745,615	4,335,930	4,409,685	0.20	0.34	0.07
16	8,776,605	4,358,460	4,418,145	8,758,551	4,343,177	4,415,374	0.21	0.35	0.06
17	8,788,436	4,365,515	4,422,921	8,770,202	4,349,852	4,420,350	0.21	0.36	0.06
18	8,798,865	4,371,910	4,426,955	8,780,511	4,355,907	4,424,604	0.21	0.37	0.05
19	8,807,884	4,377,631	4,430,253	8,789,471	4,361,327	4,428,144	0.21	0.37	0.05
20	8,815,439	4,382,644	4,432,795	8,797,052	4,366,078	4,430,974	0.21	0.38	0.04
21	8,821,590	4,386,979	4,434,611	8,803,286	4,370,177	4,433,109	0.21	0.38	0.03
22	8,826,448	4,390,678	4,435,770	8,808,247	4,373,654	4,434,593	0.21	0.39	0.03
23	8,830,104	4,393,761	4,436,343	8,812,058	4,376,557	4,435,501	0.20	0.39	0.02
24	8,832,724	4,396,297	4,436,427	8,814,891	4,378,967	4,435,924	0.20	0.39	0.01
25	8,834,481	4,398,371	4,436,110	8,816,913	4,380,960	4,435,953	0.20	0.40	0.00
26	8,835,528	4,400,046	4,435,482	8,818,248	4,382,582	4,435,666	0.20	0.40	0.00
27	8,836,099	4,401,420	4,434,679	8,819,153	4,383,939	4,435,214	0.19	0.40	-0.01
28	8,836,476	4,402,627	4,433,849	8,819,925	4,385,175	4,434,750	0.19	0.40	-0.02
29	8,836,913	4,403,798	4,433,115	8,820,760	4,386,382	4,434,378	0.18	0.40	-0.03
30	8,837,666	4,405,051	4,432,615	8,821,905	4,387,664	4,434,241	0.18	0.39	-0.04

3 Stochastic Variation in Mortality

3.1 Embedding natural variation in the projections

Figs. 3.1-3.4 show estimates of α for 1950-70 and 1980-2002 for males and females, respectively. In the case of 1950-70, 1950 survival is the standard. In the case of 1980-2002, it is 1980 survival that serves as standard. It will be noted that the time-patterns are nearly linear for which reason the corresponding regression lines have been shown. The time-patterns in α for the Danish life tables (table 1.1) bring forth similar linear features. This is an advantageous feature of (1.4) since generally changes in life expectancy become linear changes in α.

Fig. 3.1. Time progression for alpha, Swedish males 1950-1970

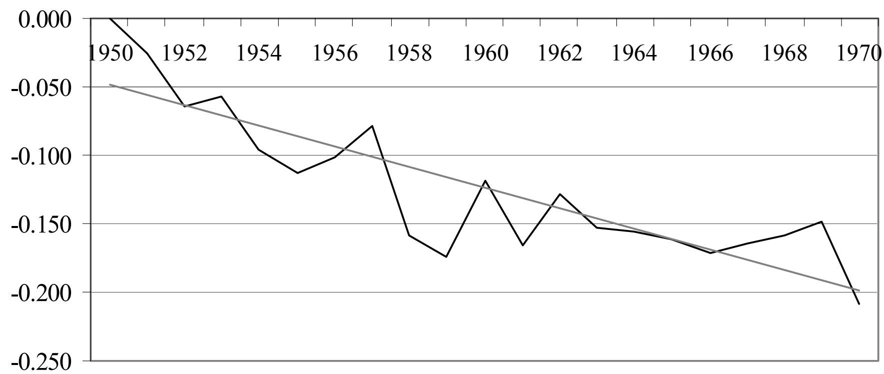


Fig. 3.2. Time progression for alpha, Swedish females 1950-70

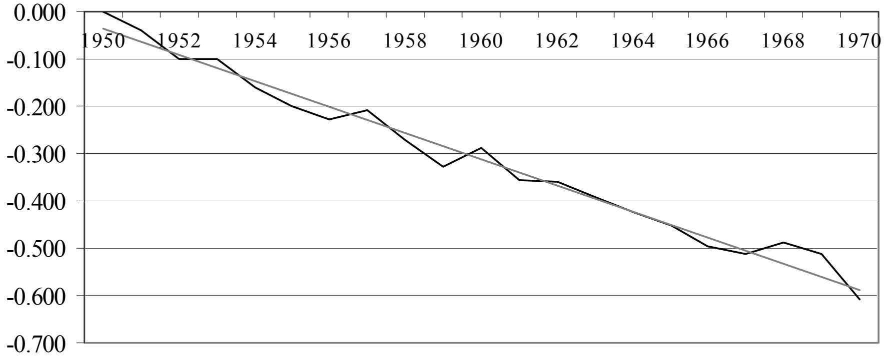


Fig. 3.3. Time progression for alpha, Swedish males 1980-2002

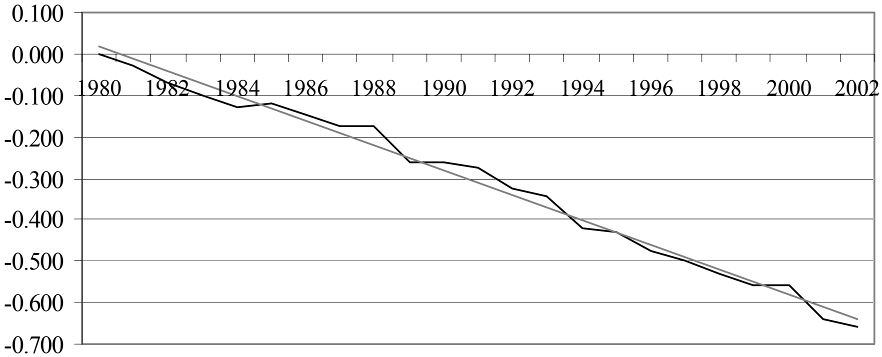
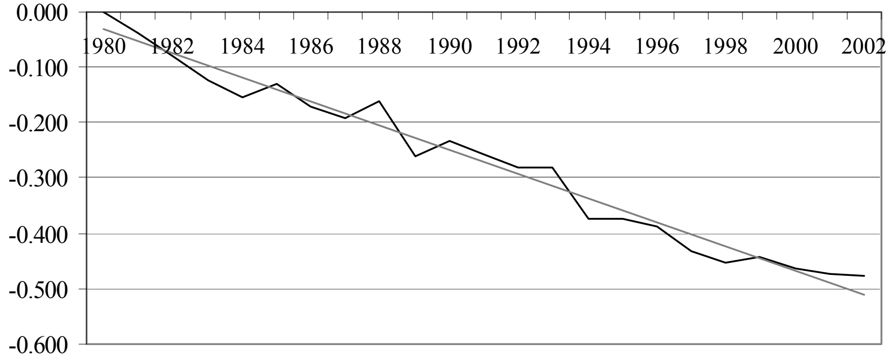


Fig. 3.4. Time progression for alpha, females 1980-2002



It seems reasonable, then, to model the time progression of α_t as

$$\alpha_t = h + kt + e_t,$$

that is, as a linear trend plus white noise.

For the period 1950-70, we find for males

$$\alpha_t = -0.0492 - 0.0075t + e_t,\ \sigma_e^m = 0.028 \tag{3.1}$$

and for females

$$\alpha_t = -0.0355 - 0.0275t + e_t,\ \sigma_e^f = 0.025 \tag{3.2}$$

$t = 0, \dots, 20.$

For the period 1980-2002, we find for males

$$\alpha_t = 0.0162 - 0.03t + e_t,\ \sigma_e^m = 0.023 \tag{3.3}$$

and for females

$$\alpha_t = -0.031 - 0.022t + e_t,\ \sigma_e^f = 0.025 \tag{3.4}$$

$t = 0, \dots, 22.$

It can be shown that for (3.1) – (3.4) residuals perform like white noise with the indicated standard deviations. It will be seen that for all four representations, $\sigma_e \approx 0.03$. For illustrative purposes, fig. 3.5 and fig. 3.6 show the virtually linear relationships between life expectancy and α for Sweden 1980-2002. Table 3.1 summarizes estimated regression representations.

Fig. 3.5. The relationship between a and the life expectancy at birth for Swedish males, 1980-2002

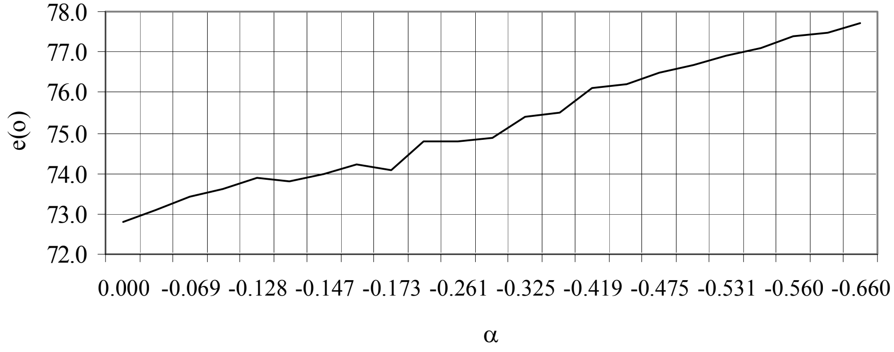


Fig. 3.6. The relationship between a and life expectancy at birth for Swedish females, 1980-2002

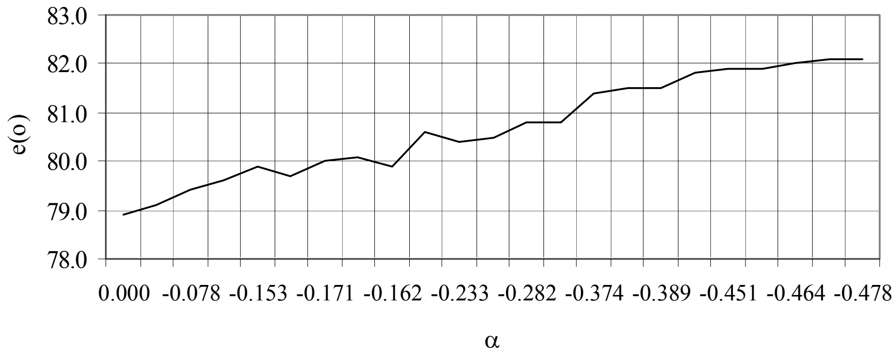


Table 3.1 Estimated coefficients for regression representations $\hat{\alpha}(t) = \hat{k}\,t + \hat{h}$, and standard errors about regression lines (white noise). Sweden 1950-70 and 1980-2002 and Denmark 1921-51

Period	$\hat{k}$	$\hat{h}$	$\hat{\sigma}_e$	t
Sweden				
Males				
1950-70	-0.0075	-0.0492	0.028	$0 \le t \le 21$
1980-2002	-0.0298	0.0162	0.023	$0 \le t \le 22$
Females				
1950-70	-0.0275	-0.0355	0.025	$0 \le t \le 21$
1980-2002	-0.0219	-0.0305	0.025	$0 \le t \le 22$
Denmark				
1921-51				
Males	-0.0216	0.136	0.064	$0 \le t \le 30$
Females	-0.0255	0.148	0.075	$0 \le t \le 30$

$\hat{\sigma}_e$ is the standard deviation for white noise.

References

Brass, W (1971). *On the Scale of Mortality* in Biological Aspects of Demography, ed. By William Brass, pp. 69-110.

Brass et al. (1968). *The Demography of Tropical Africa*. Princeton: Princeton University Press.

Brass, W (1974). Mortality Models and their Uses in Demography. *Transactions of the Faculty of Actuaries*, Vol. 33, No. 2, pp. 123-142.

Brass, W (1975). *Methods for Estimating Fertility and Mortality from Limited and Defective Data.* The Carolina Population Center. University of North Carolina at Chapell Hill.

Carrier, N. H. & Goh, T.J (1972). The Validation of Brass's Model Life Table System. *Population Studies*, Vol. 26, No. 1, pp. 29-51.

Mortality model 2:

Old Age Mortality in Stochastic Population Projections[10]
Michael Hartmann
Statistics Sweden
October 2005

Abstract

Past increases in longevity were mainly attributable to falling infant and childhood mortality. It is now contemplated that further increases in longevity must come from falling adult and old age mortality. This paper deals with mortality in the context of stochastic population projections and addresses a scenario where changes in mortality are limited to take place at ages 55 and over. Specifically, it is shown that for Sweden during the 20th century, the Brass logit survival model provides a parsimonious description of survival after age 55. It is demonstrated how time series of model-parameters can be used to predict future stochastic survival functions.

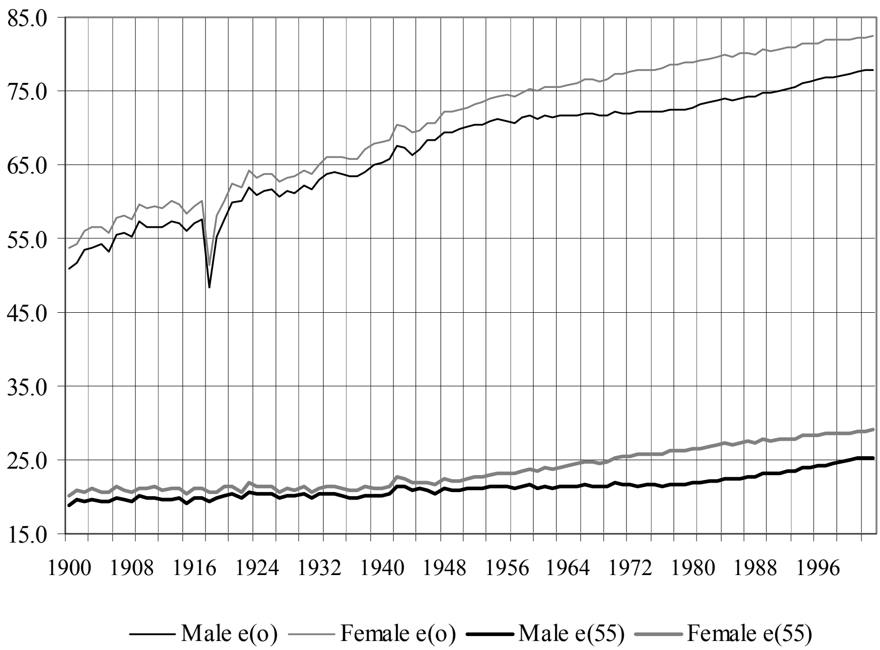
[10] Paper presented at "Symposium i anvendt statistik", Informatik og Matematisk Modellering, Danmarks Tekniske Universitet, Danmarks Statistik, 23-25 January, 2006. Published in proceedings.

1 Introduction

1.1 The need for projecting old age mortality

Although mortality usually plays a minor role in ordinary population projections, especially if they are terminated at ages 75 or 80 years (common ages of termination in many population projection packages), mortality plays a more salient role when projecting the population beyond these ages. In addition, it must be noted that in countries with high life expectancies, about fifty percent of deaths take place after age 75.

Fig. 1.1. Life expectancies at birth and at age 55 by sex for Sweden 1900-2003



It was not until the beginning of the 1940s that significant improvements in survival above age 55 began to take place (fig. 1.1); improvements in old age mortality is a typical post World War II demographic feature of industrialized societies. In passing, it will be noted that increases in male life expectancies at birth during the 1950s and 1960s were slower than for females but that at the beginning of the 1970s male life expectancies began to improve as rapidly as for females, albeit at a lower level. Also, while the remaining life expectancy at age 55 began to increase markedly for females toward the end of the 1940s, similar increases for males did not begin to take place until the early 1970s (fig. 1.1). In this paper we have chosen age 55 as a breaking point between young and old mortality.

1.2 Projection probabilities

The specification of mortality in the case of population projections involves the survival function and its corresponding person-years (exposure times). In ordinary life table notation, the survival function l_x gives the probability for an individual to survive from birth until age x. Expressed in life table mortality rates q_x (the probability for a life aged x to die before reaching age x+1), the survival function is

$$l_x = \prod_{o}^{x-1}(1-q_j) \tag{1.1}$$

The person-years lived by individuals at age x is

$$L(x) = \int_{x}^{x+1} l_t \, dt$$

which usually are approximated by

$$L(x) \approx (l_x + l_{x+1})/2$$

The probability projecting the population from age x to age x+1 is

$$\pi_x = L(x+1)/L(x) \tag{1.2}$$

In the context of population projections, survival of the population from one age to the next is determined by the invoked survival functions and their associated projection probabilities (1.2) for males and females, respectively, with allowance for time-dependence. Because the survival function (1.1) is a product of survival rates ($p_x = 1 - q_x$), different sets of mortality rates may yield almost indistinguishable projection probabilities (1.2). This, among other things, is one of the reasons why projected age-distributions, especially if they are terminated at standard ages such as 75 or 80 years, often are relatively insensitive to the choice of survival function.

2 The Brass Logit Survival Model

2.1 A pseudo-logistic relationship

Life table survival functions can be related in the sense that for a chosen standard survival function l_x^s another survival function l_x can be expressed as

$$l_x \approx 1/(1+e^{\alpha}((1-l_x^s)/l_x^s)^{\beta}) \tag{2.1}$$

where α and β are parameters (Brass, 1971, 1974; Hill and Trussell, 1977). Denoting estimated parameters by $\hat{\alpha}$ and $\hat{\beta}$, modeled survival is

$$\hat{l}_x(\hat{\alpha},\hat{\beta}) = 1/(1+e^{\hat{\alpha}}((1-l_x^s)/l_x^s)^{\hat{\beta}}) \tag{2.2}$$

The significance of the parameters is that, relative to chosen standard survival, α is related to the level of mortality while β controls the relationship between childhood and adult mortality (Brass, 1974). It will be noted that (2.1) implies that

$$\text{logit}\, l_x = \alpha + \beta\, \text{logit}\, l_x^s \tag{2.3}$$

with $\text{logit}\, p = \ln\frac{1-p}{p}$, $0 < p < 1$. Letting $l_x = 1 - F(x)$, and assuming that $F_s(x)$ in $l_x^s = 1 - F_s(x)$ is a logistic distribution with mean μ_s and variance σ_s^2, the modeled distribution function F(x) is logistic with mean $\mu_s - \alpha$ and variance $\beta^2\sigma_s^2$. Hence, e.g., for $\alpha < 0$, the mean age at death for modeled survival is increased (relative to the standard). When the scale-parameter $\beta = 1$, the variance of the distribution of modeled deaths is the same as for the standard distribution of deaths. If e.g., $\beta < 1$, the variance of modeled deaths is less than for standard distribution of deaths. In reality, the distribution of deaths is not logistic. However by treating it as if though it were, the pseudo-logistic relationship (2.3) emerges with the result that α measures the level of mortality of l_x and β the variance of the distribution of deaths of l_x relative to l^s_x, respectively. As noted, in this paper survival begins at age 55, that is, x is replaced by x-55, $x \geq 55$ in (2.1).

2.2 Survival above age 55

As noted, there was steady progress in longevity during the 20th century (except during the years of the influenza pandemic, c. 1916-17). Nevertheless, survival among individuals aged 55 did not begin to improve materially until the 1940s. It is now widely believed that future survival improvements will take place at adult and late adult ages. Specifically, for the purpose of projecting the Swedish population, it was contemplated that, during the projection period, mortality at ages below 55 would be more or less the same as around the year 2000 while marked future changes would be limited to ages above 55.

Consider the case when 1900 male survival is used as a standard in (2.1). In this event, the standard can be used to generate model survival for the years

1901, … , 2003 (in this paper, we work with life tables for the period 1900-2003). This yields parametric time series $\hat{\alpha}_t$ and $\hat{\beta}_t$, t = 1900, … , 2003.

Fig. 2.1 shows estimates of $\hat{\alpha}_t$ when using male standard survival for the years 1900, 1910, 1920 and 1930. It will be seen that the four standards lead to parallel time-patterns. Fig. 2.2 shows corresponding estimates when using male standard survival for the years 1940, 1950, 1960, 1970 and 1980. Figs. 2.3-2.8 display the time-patterns of the parameters for the remaining male and female standards.

Fig. 2.1. Estimates of alpha using standard survival for 1900, 1910, 1920 and 1930: Swedish males

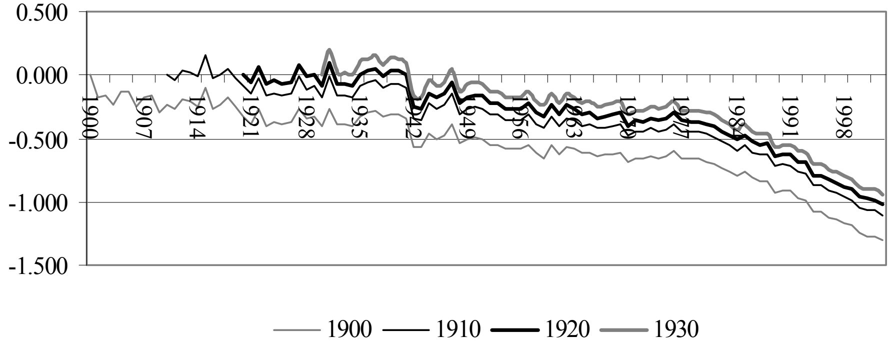


Fig. 2.2. Estimates of alpha using standard survival for 1940, 1950, 1960, 1970 and 1980: Swedish males

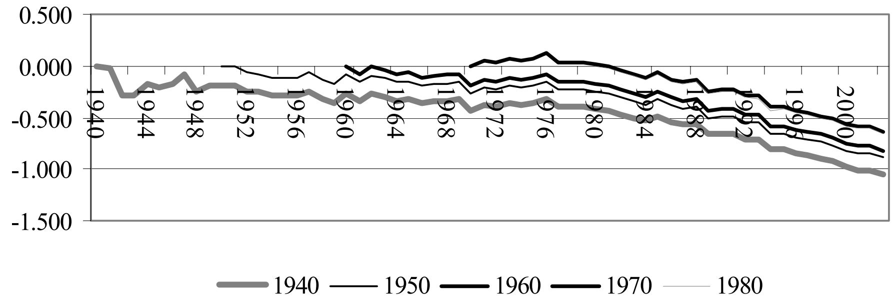


Fig. 2.3. Estimates of beta using standard survival for 1900, 1910, 1920, and 1930: Swedish males

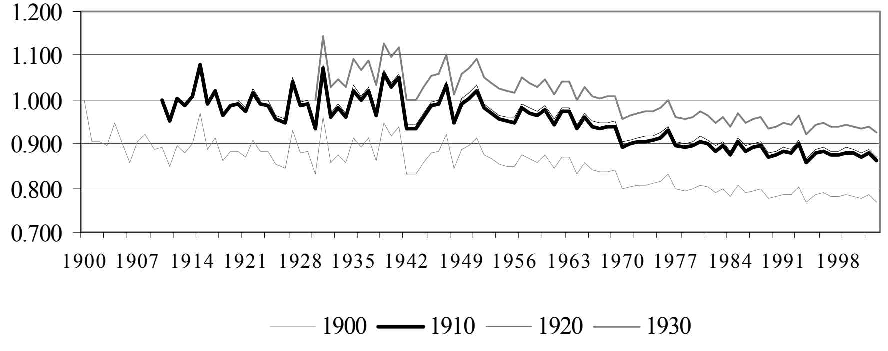


Fig. 2.4. Estimates of beta using standard survival for 1940, 1950, 1960, 1970 and 1980: Swedish males

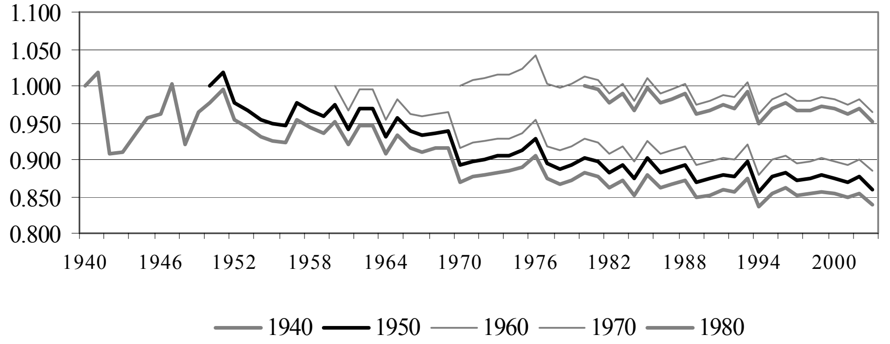


Fig. 2.5. Estimates of alpha using standard survival for 1900, 1910, 1920 and 1930: Swedish females

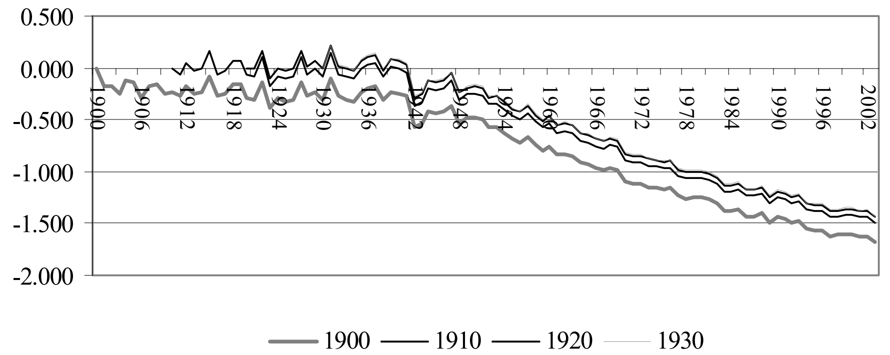


Fig. 2.6. Estimates of alpha using standard survival for 1940, 1950, 1960, 1970 and 1980: Swedish females

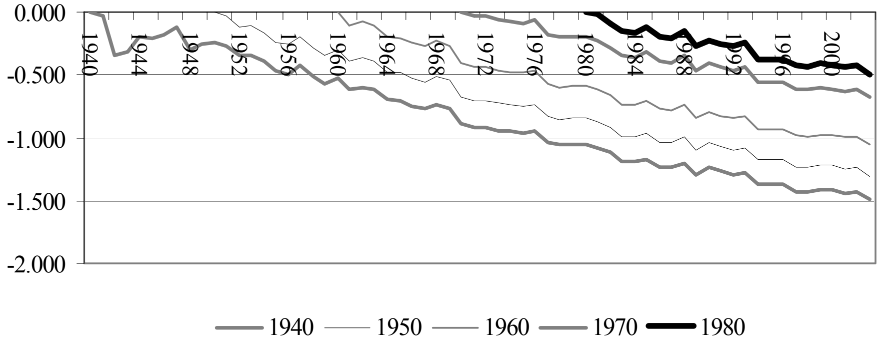


Fig. 2.7. Estimates of beta using standard survival for 1900, 1910, 1920, and 1930: Swedish females

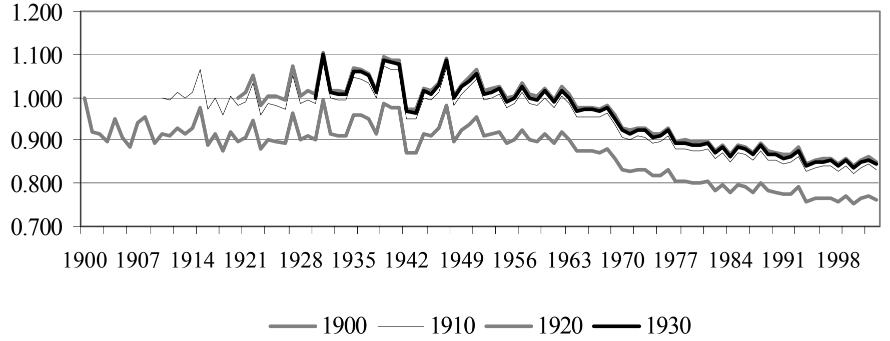
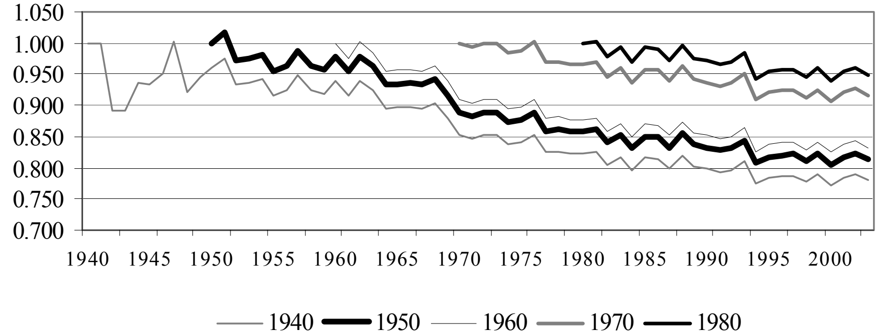


Fig. 2.8. Estimates of beta using standard survival for 1940, 1950, 1960, 1970 and 1980: Swedish females



There is considerable regularity in figs. 2.1-2.8. Regardless of which standard is used, the time-patterns for $\hat{\alpha}_t$ and $\hat{\beta}_t$ are nearly the same, although they unfold at different levels. For males it is possible to distinguish three different periods of survival behavior. Between 1900 and the early 1940s, there is relative stability in the sense that the level of mortality, as measured by $\hat{\alpha}_t$, hardly changes (figs. 2.1 and 2.3). Between the 1940s and early 1970s there is a beginning decline in $\hat{\alpha}_t$. After the beginning of the 1970s, there is a steep decline in $\hat{\alpha}_t$ pointing to a marked increase in remaining life expectancy at age 55. Especially for 1970 and 1980 standard survival, there are steady and similar time-patterns in $\hat{\alpha}_t$. In the case of $\hat{\beta}_t$ (figs. 2.2 and 2.4) there is a tendency for $\hat{\beta}_t$ to drop between 1900 and the early 1970s. After 1976 there is a marked drop in $\hat{\beta}_t$ pointing to a "narrowing variance" in the distribution of deaths.

For females (figs. 2.5 and 2.6), estimates of $\hat{\alpha}_t$ retain almost constant levels until the beginning of the 1940s. After the beginning of the 1940s there is a marked fall in $\hat{\alpha}_t$ pointing to a steady increase in remaining life expectancy at age 55. With respect to $\hat{\beta}_t$ they hover around almost constant levels until the beginning of the 1940s after which they drop systematically. Especially for 1970 and 1980 standard survival, there is a steady decrease in $\hat{\beta}_t$. Stated differently, the survival model (2.1), by means of its level and scale parameters, captures important temporal details of the age-pattern of mortality.

As an example of the close fit provided by (2.1) using male survival for the year 1980 as a standard, observed and modeled survival functions for the year 2000 are shown in fig. 2.8 (the fits for males and females are the same and the observed and modeled curves almost coincide). For males, e(55) in the observed table is 26.0 and for the modeled 26.1 years; a minuscule difference considering that the standard dates back twenty years earlier. Fig. 2.10 shows the age-distribution of deaths for the observed and modeled 2000 male survival functions (using the 1980 male survival function as a standard).

Fig. 2.9. Observed and modeled survival functions for the year 2000 using the 1980 survival function as standard: Swedish males

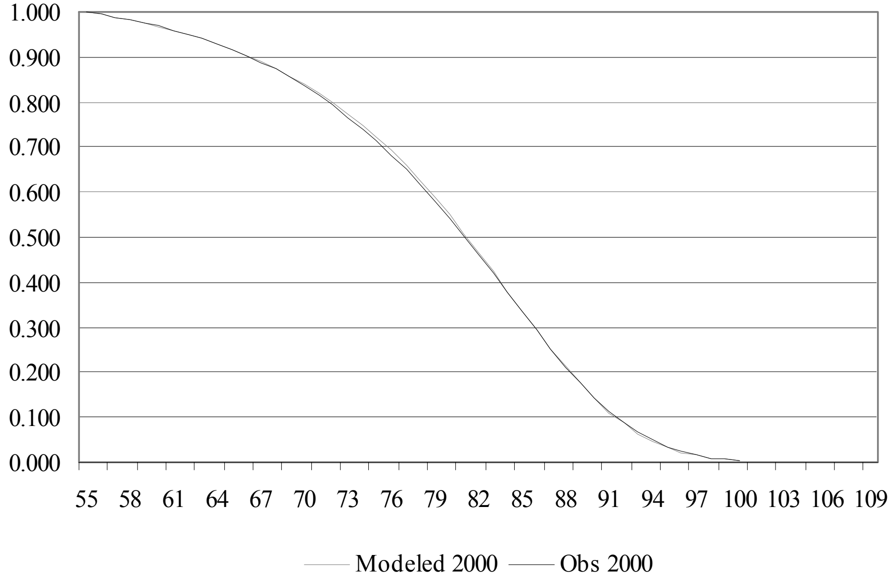
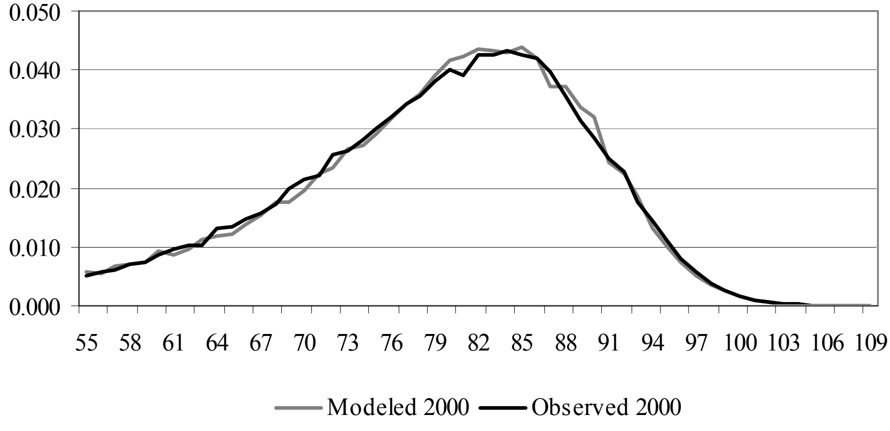


Fig. 2.10. The distribution of deaths at ages 55 and over for modeled and observed male survival functions in year 2000, using 1980 male survival as standard



3 Extrapolating Survival to Year 2030

3.1 Regression representations using 1980 standard survival

The main features of the parameter trajectories for males and females, relative to1980 standard survival, are somewhat different. Whereas for males there is a clear linear downward trend in $\hat{\alpha}_t$, the trend for females appears more like a logarithmic curve. This is illustrated by figs. 3.1 and 3.2, respectively. For both males and females, the decline in β_t is very nearly linear. Regression representations for α_t and β_t, when using 1980 survival as standards, are given in table 3.1.

Fig. 3.1. Extrapolation of alpha to year 2030 using 1980 Swedish male survival as standard

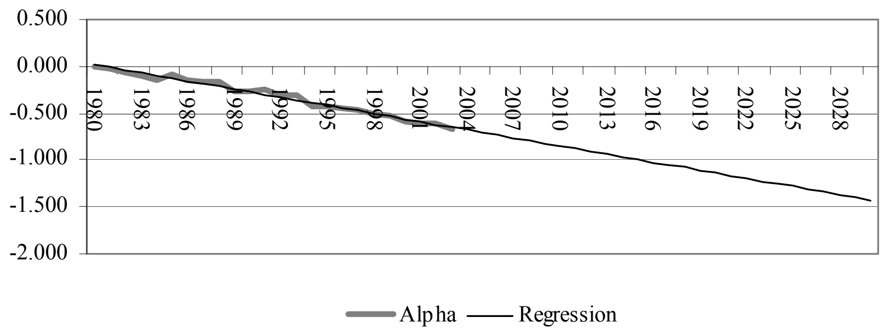


Fig. 3.2. Extrapolation of alpha to year 2030 using 1980 Swedish female survival as standard

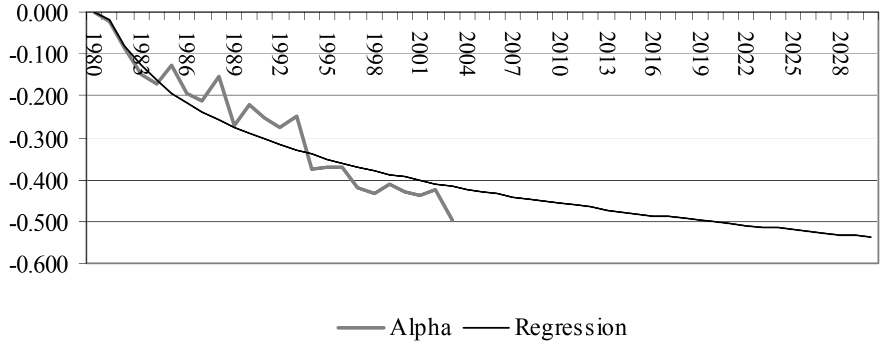


Fig. 3.3. Extrapolation of beta to year 2030 using 1980 Swedish male survival as standard

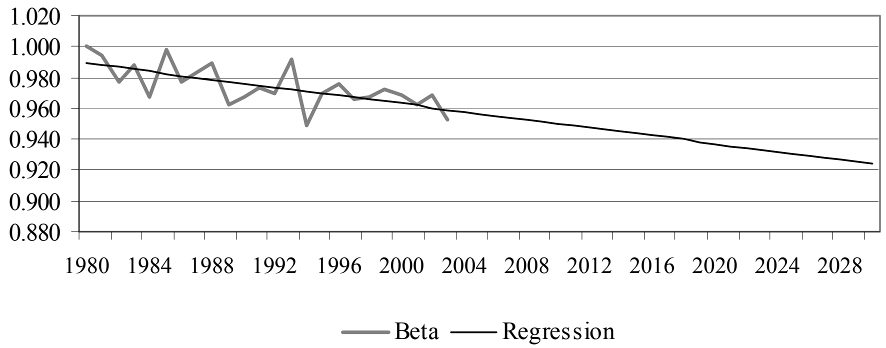


Fig. 3.4. Extrapolation of beta to year 2030 using 1980 Swedish female survival as standard

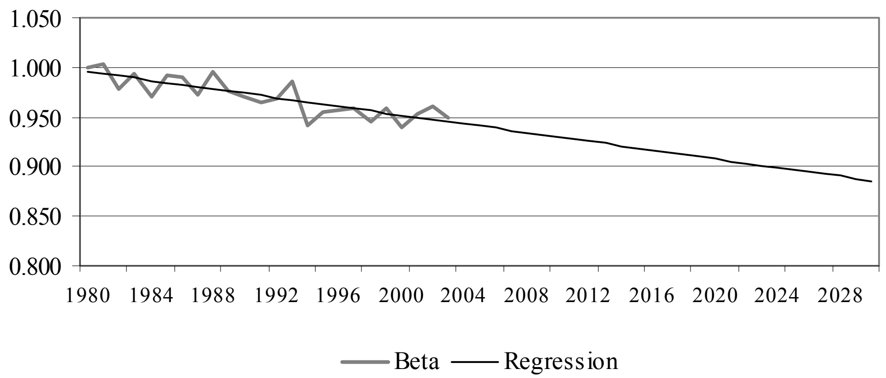


Table 3.1. Regression representations for α_t and β_t extending to the year 2030, using 1980 survival as standards for males and females

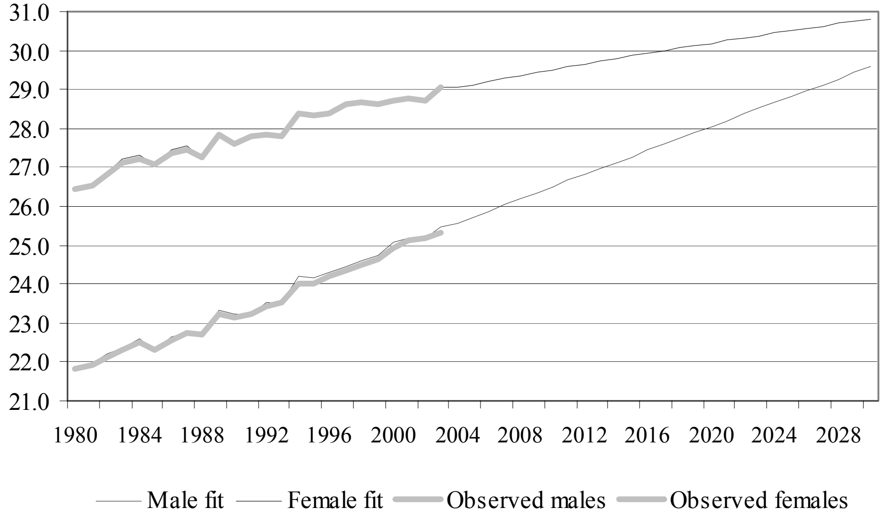
Sex	1980 standard survival	Error $\hat{\sigma}_e$
	Regression for α_t	
Males	-0.0289 (t-1979) +0.045	0.03
Females	-0.160 ln (t-1979) + 0.094	0.05
	Regression for β_t	
Males	-0.0013 (t-1979) + 0.990	0.01
Females	-0.0022 (t-1979) + 0.998	0.01

Extrapolation of any time series involves uncertainty. Especially if the extrapolation extends far into the future, there is reason to believe that during its course it will come to deviate from the actual unfolding of the process; this is the omnipresent problem of extrapolation or forecasting.

The parameter extrapolations given in table 3.1 and in figs. 3.1-3.4 reflect the assumption that mortality above age 55 will decline with the same tempo as during the past 20 years or so.

Fig. 3.5 shows modeled life expectancies at age 55 for males and females based on the regression representation in table 3.1 (these regression representations are derived from 1980 standard survival for males and females, respectively). For comparison, fig. 3.5 also shows the observed life expectancies at age 55 for males and females for the period 1980-2003. It will be noted that modeled expectancies virtually coincide with the observed ones during 1980-2003. This shows, at least for the period 1980-2003, that modeled expectancies derived from the 1980 standard maintain high precision, even some twenty years later than the reference period of the standard. Table 3.2 shows modeled and the observed expectancies shown in fig. 3.5.

Fig. 3.5. Observed and modeled remaining life expectancies at age 55 for Swedish males and females



The regression representations for α_t and β_t eventually lead to a remaining life expectancy at age 55 of 29.6 for males and 30.8 years for females. Stated differently, if the improvements in mortality after age 55 observed during the period 1980-2003 were to continue with the same tempo of parametric time-patterns, then males would achieve a remaining life expectancy at age 55 of about 30 years and females would achieve a corresponding expectancy of 31 years in the year 2030. There is, then, a projected "narrowing gap" between male and female expectancies at age 55 during the extrapolation period.

3.2 Stochastic extrapolation of the remaining life expectancy at age 55

If α and β in (2.1) are replaced by time series α_t and β_t, respectively,

$$s(t;x) = \frac{1}{1 + e^{\alpha_t} \left[\frac{1 - l_x^s}{l_x^s} \right]^{\beta_t}} \tag{3.1}$$

becomes a random function.

The close fit provided by (2.1) during long periods when $x \geq 55$ cannot be achieved for $x \geq 0$; there is, understandably enough, more regularity in the age-pattern of survival curves at ages above 55 years than over the entire age-span. This also suggests that by application of a standard survival curve that references a period shortly before the onset of the projection period, this standard, once it is embedded in (2.1) or (3.1), is likely to provide a parsimonious description of future survival, insofar as the projection period is not too long.

Fig. 3.6. Simulated life expectancies at age 55 for males and females (based on 1980 standard survival)

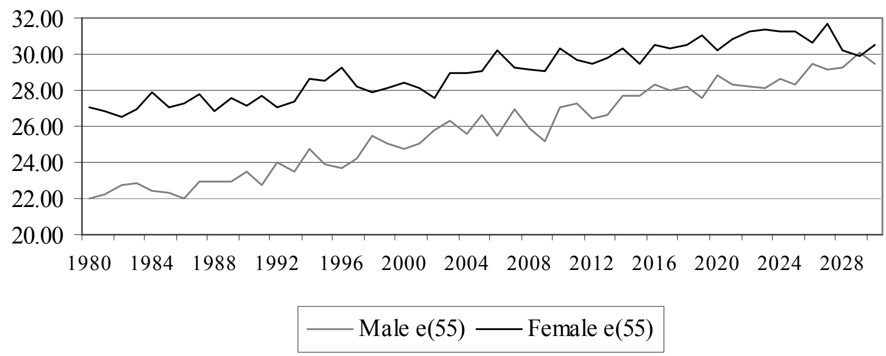
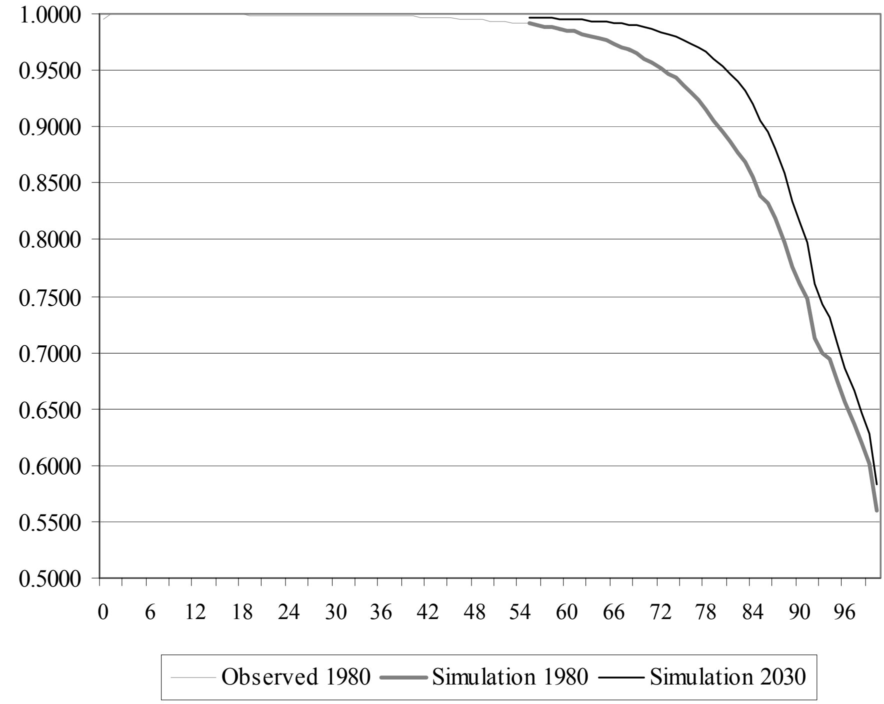


Fig. 3.6 shows simulations of the life expectancies at age 55 for males and females, respectively (1980 standard survival). The standard deviations of error terms (independent and normally distributed) in the parametric regressions are those indicated in table 3.1. It will be noted that for these realizations the life expectancies at age 55 for males and females cross over in the year 2029.

Fig. 3.7. Projection probabilities for males in 1980 and 2030. Probabilities for ages less than 55 are those of the 1980 male survival function. Simulated values begin at age 55

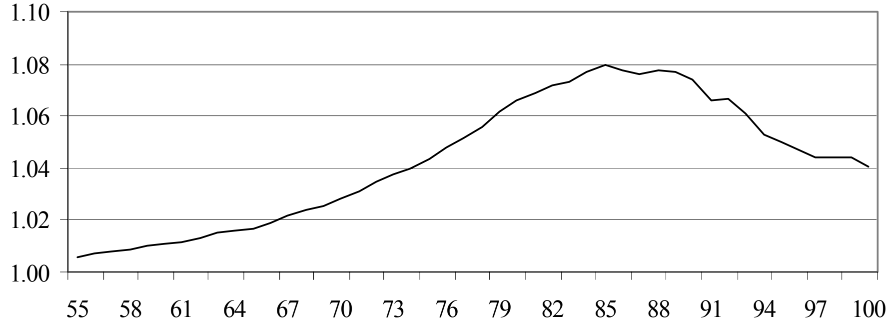


A projection period of thirty or more years is very long and for this reason is mainly supportive of academic discussion. Nevertheless, to get an impression of likely future variation in mortality, given that variability of the past will more or less repeat itself, (3.1) is a suitable survival model, especially since the projection probabilities π_x in (1.2) come from the survival function, rather than from the mortality rates q_x (or central death rates m_x).

Fig. 3.7 shows the projection probabilities (1.2) for males in 1980 and 2030. The curve reflects that the projection probabilities at ages below 55 remain those of the 1980 survival function for males while the projection probabilities at ages 55 and over are simulated by means of the regression models in table 3.1. As in the case of fig. 3.6 the error terms in the regression models are assumed independent and normally distributed with zero mean and the standard deviations given in table 3.1. The small undulations in the curves result from the 1980 survival standard not having been smoothed. In should be noted though that such smoothing effaces in an actual projection of the population.

The effect of the model (3.1) is that it leads to an improvement in survival (relative to the 1980 standard survival function for males) that increases up to about age 85. After age 85 there is still some improvement, but it declines rapidly toward the end of life (fig. 3.8).

Fig. 3.8. Ratios between simulated projection probabilities for males in 2030 and 1980 (1980 standard survival)



4 Using Recent Standard Survival

4.1 The life table for 2000

To project the Swedish population thirty years into the future, it might seem reasonable to draw on a survival age-pattern the timing of which is close to the beginning of the projection period, rather than on 1980 survival. Generally, a recently observed survival function can be used as a standard for modeling past survival. The resulting time series $\hat{\alpha}_t$ and $\hat{\beta}_t$ can be extrapolated, as previously demonstrated.

Because extrapolation of stochastic processes, whether stationary or not, always is ridden with uncertainty, and sometimes considerable, it is important not to see some recent trend as an indication of survival behavior many years into the future. To this must be added that, generally, it may be difficult to distinguish between a "trend" and mere stationary random behavior. As noted, the hypothesis that lies behind extrapolation in the present paper is that survival improvements primarily will take place at ages above 55; the view being that there is little latitude for improving survival below this age.

Fig. 4.1. Time-pattern of alpha, observed and extrapolated, for males when using 2000 standard survival

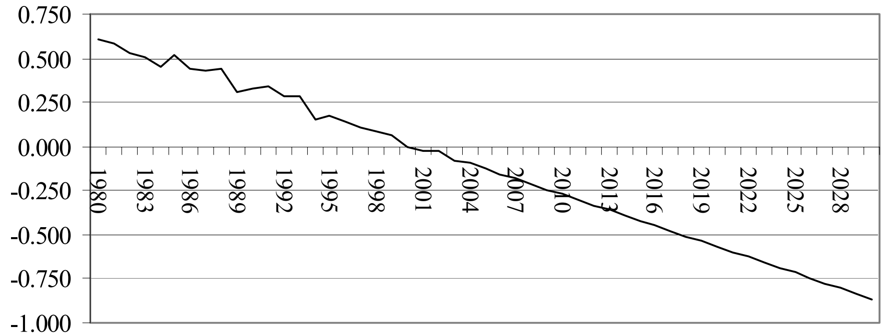


Fig. 4.2. Time-pattern of beta, observed and extrapolated, for males when using 2000 standard survival

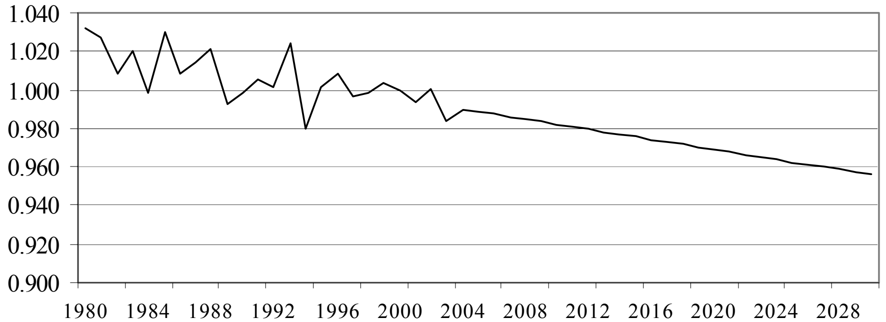


Fig. 4.3. Time-pattern of alpha, observed and extrapolated, for females when using 2000 standard survival

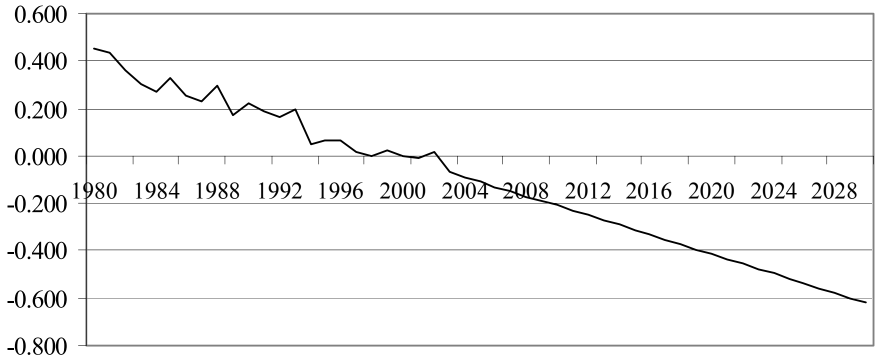


Fig. 4.4. Time-pattern of beta, observed and extrapolated, for females when using 2000 standard survival

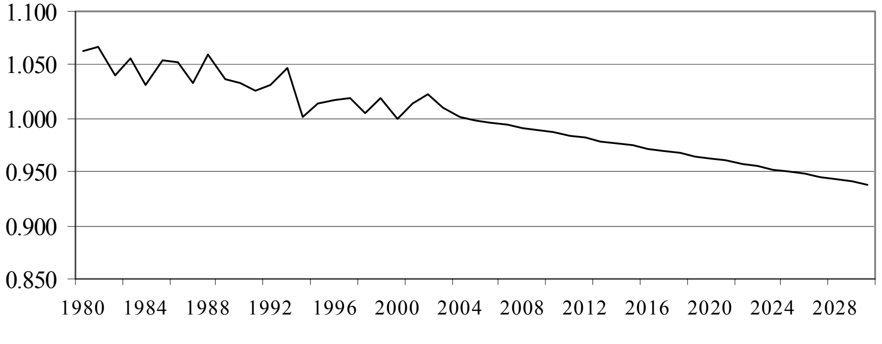


Table 4.1. Regression representations for α_t and β_t extending to the year 2030 using 2000 survival as standards for males and females

Sex	2000 standard survival	Error $\hat{\sigma}_e$
	Regression for α_t	
Males	-0.0296 (t-1979) + 0.647	0.03
Females	-0.0205 (t-1979) + 0.424	0.05
	Regression for β_t	
Males	-0.0013 (t-1979) + 1.022	0.01
Females	-0.0024 (t-1979) + 1.061	0.01

Figs. 4.1-4.4 show the parametric time-patterns for $\hat{\alpha}_t$ and $\hat{\beta}_t$ for males and females, respectively, when using 2000 standard survival. The estimated regression representations are given in table 4.1. These are similar to the ones given in table 3.1. Variation due to regression is of the same magnitude as in the case of table 3.1.

Fig. 4.5. Observed survival in 2000 and modeled survival in 2030 for Swedish males

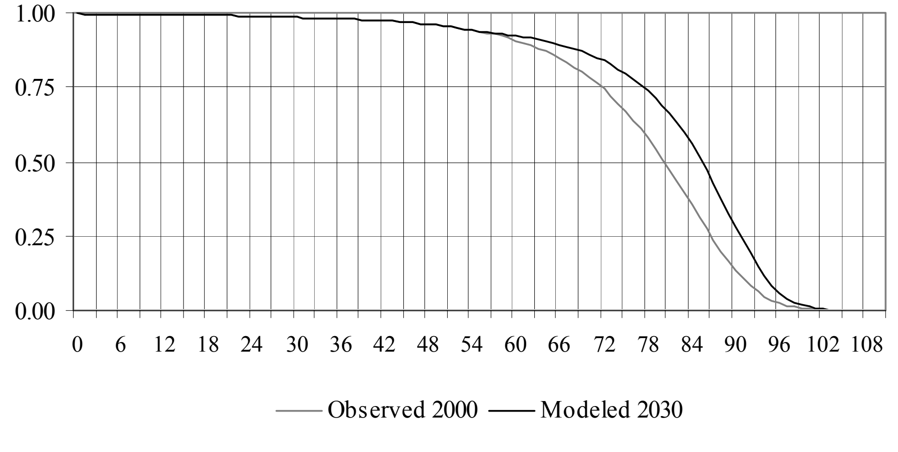


Fig. 4.5 shows the survival function for males in 2000 (the standard) and estimated male survival in 2030. Estimated survival for 2030 is such that for all ages below 55, the survival rates are the same as for 2000. Only at ages 55 and over are the survival rates changed. Fig. 4.6 shows the corresponding survival curves for females. Fig. 4.7

shows simulated life expectancies at birth for males and females, 2000-2030. These simulations are based on the error terms in table 4.1 (normally distributed errors). Fig. 4.8 shows the corresponding simulated life expectancies at age 55 for males and females. Fig. 4.9 shows simulated projection probabilities for males and females in 2000 and 2030.

Fig. 4.6. Observed survival in 2000 and modeled survival in 2030 for Swedish females

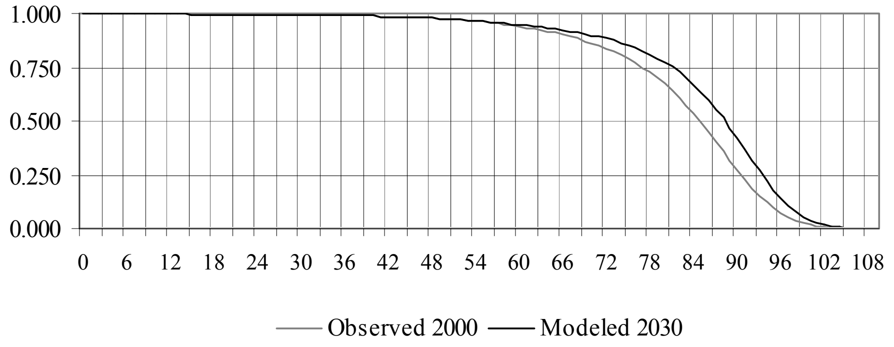


Fig. 4.7. Simulated life expectancies at birth for males and females, 2000-2030

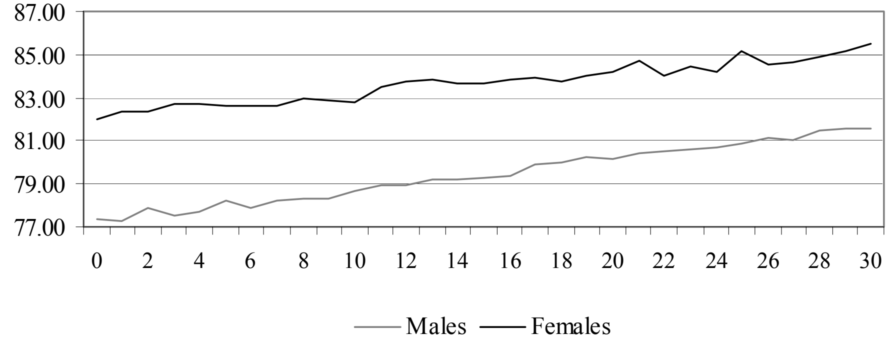
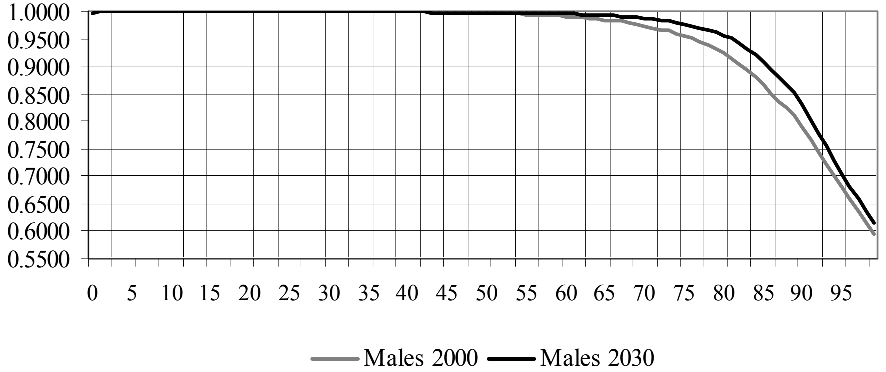


Fig. 4.8. Simulated life expectancies at age 55 for males and females, 2000-2030

Fig. 4.9. Simulated projection probabilities for males, 2000 and 2030



References

Brass, W (1971). *On the Scale of Mortality* in Biological Aspects of Demography, ed. By William Brass, pp. 69-110.

Brass et al. (1968). *The Demography of Tropical Africa*. Princeton: Princeton University Press.

Brass, W (1974). Mortality Models and their Uses in Demography. *Transactions of the Faculty of Actuaries*, Vol. 33, No. 2, pp. 123-142.

Brass, W (1975). *Methods for Estimating Fertility and Mortality from Limited and Defective Data.* The Carolina Population Center. University of North Carolina at Chapell Hill.

Carrier, N. H. & Goh, T.J (1972). The Validation of Brass's Model Life Table System. *Population Studies*, Vol. 26, No. 1, pp. 29-51.

Hill, K. & Trussell, J (1977). Further Developments in Indirect Mortality Estimation. *Population Studies*, Vol. 31, No. 2, pp. 313-334.

An Exponential Model with Three Parameters[11]

Gustaf Strandell
Statistics Sweden
December, 2005

1 Introduction

The number of children born in the future depends on the present as well as the future age-distribution of women. Hence, when forecasting the number of births it is necessary to take the prevailing age-distributions of women into account. Let $f_t(x)$ denote age-specific fertility rates for women at ages $x = 14, \ldots, 49$ during year t. Then the total fertility rate (TFR) for year t is

$$\mathrm{TFR}_t = \sum_{t=14}^{49} f_t(x),$$

that is, TFR_t is the number of children a women would give birth to if the age-specific fertility rates in year t would last during her entire fertile period.

[11] Paper presented at "Symposium i anvendt statistik", Informatik og Matematisk Modellering, Danmarks Tekniske Universitet, Danmarks Statistik, 23-25 January, 2006.

Fig. 1.1. Time-pattern for total fertility rate: Sweden 1900-1999

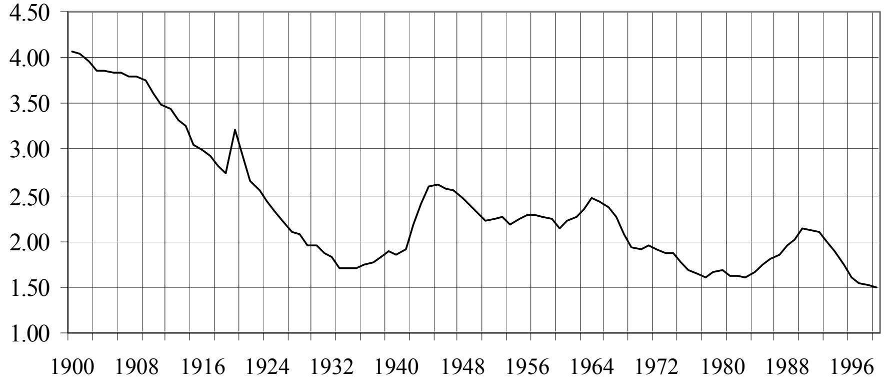


Fig 1.1 shows TFR for Sweden during the 20th century. Apart from a peak around 1920 it has followed a more or less linear decline from about TFR = 4.1 to TFR = 1.7 between 1900 and 1933. Since then TFR has fluctuated in a band extending from a floor of TFR = 1.5 (a few years before the turn of the 20th century) to a ceiling of 2.6 (in 1945). The mean for 1933-2000 is TFR = 2.0.

A simple and popular method used to predict future age-specific fertility rates consists of keeping the width and the central location of the fertility curve fixed during the projection period, allowing only the TFR to change. In Section 2 it is demonstrated that in Sweden this approach leads to systematic errors. As an alternative, we introduce in Section 3 an exponential model of age-specific fertility which, besides TFR, uses two other important characteristics of the fertility curve as parameters. In Section 4 we look closer at how these parameters have behaved for Sweden between 1970 and 2000. We also estimate models for observed time series and use these to extrapolate the parameters into the future.

2A One-Parameter Model of Fertility

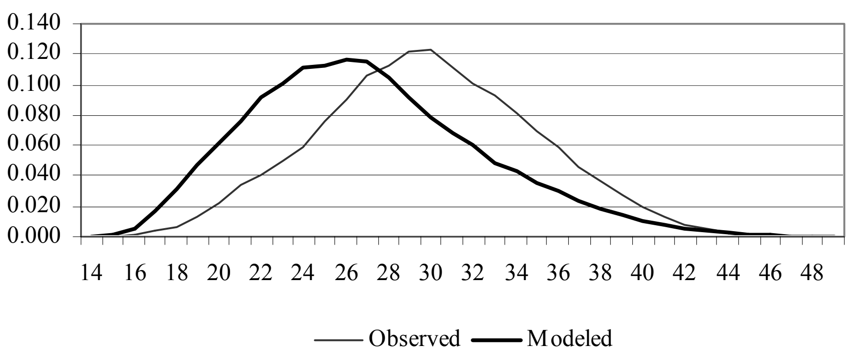
A simple method that is sometimes used to forecast age-specific fertility rates consists of keeping the age-pattern of the fertility schedule fixed and thus only predict the future level of TFR. Basing such a prediction on age-specific fertility rates observed in the year 1970, we would predict fertility rates

$f_{x,t}$ in year t by

$$f_{x,t} = \frac{TFR_t}{TFR_{1970}} f_{x,1970} \qquad 14 \le x \le 49 \tag{2.1}$$

Fig. 2.1 shows predicted age-specific fertility rates for the year 2000 using (2.1) as well as observed fertility rates for 2000.

Fig. 2.1. Observed and modeled age-specific fertility rates for Sweden, 2000



As will be seen, observed and forecasted rates do not agree very well. Applying these two sets of fertility rates to the 2000 mid-year population of women gives the number of children shown in table 2.1.

Table 1. Observed and estimated number of births in 2000

Observed	Estimated	Error	Percent error
90,441	86,770	3.67	4.1

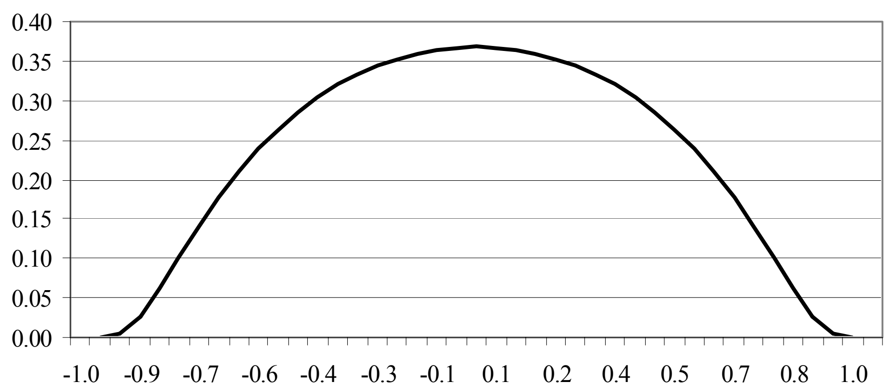
3A Three-Parameter Model of Age-Specific Fertility

In this section we introduce a simple model that describes the time evolution of age-specific fertility rates using three parameters. Based on the findings in the last section, in addition to TFR, we will add a parameter that controls the placement of the top of the fertility curve. We will also add a parameter that controls the width of the top. The inspiration for the choice of model comes from *distribution theory*; a branch of mathematics dealing with partial differential equations (Hörmander, 1963). At the base of distribution theory is the notion of a *test-function*. A test-function on the real line is a function which is infinitely differentiable and zero outside a bounded subset (properties which are not important to us here). The simplest test-function is

$$t(x) = \begin{cases} e^{\frac{-1}{1-x^2}} & \text{if } -1 < x < 1 \cdot \\ 0 & \text{elsewhere} \end{cases}$$

the graph of which is shown in fig. 3.1.

Fig. 3.1. Simple test-function t(x)



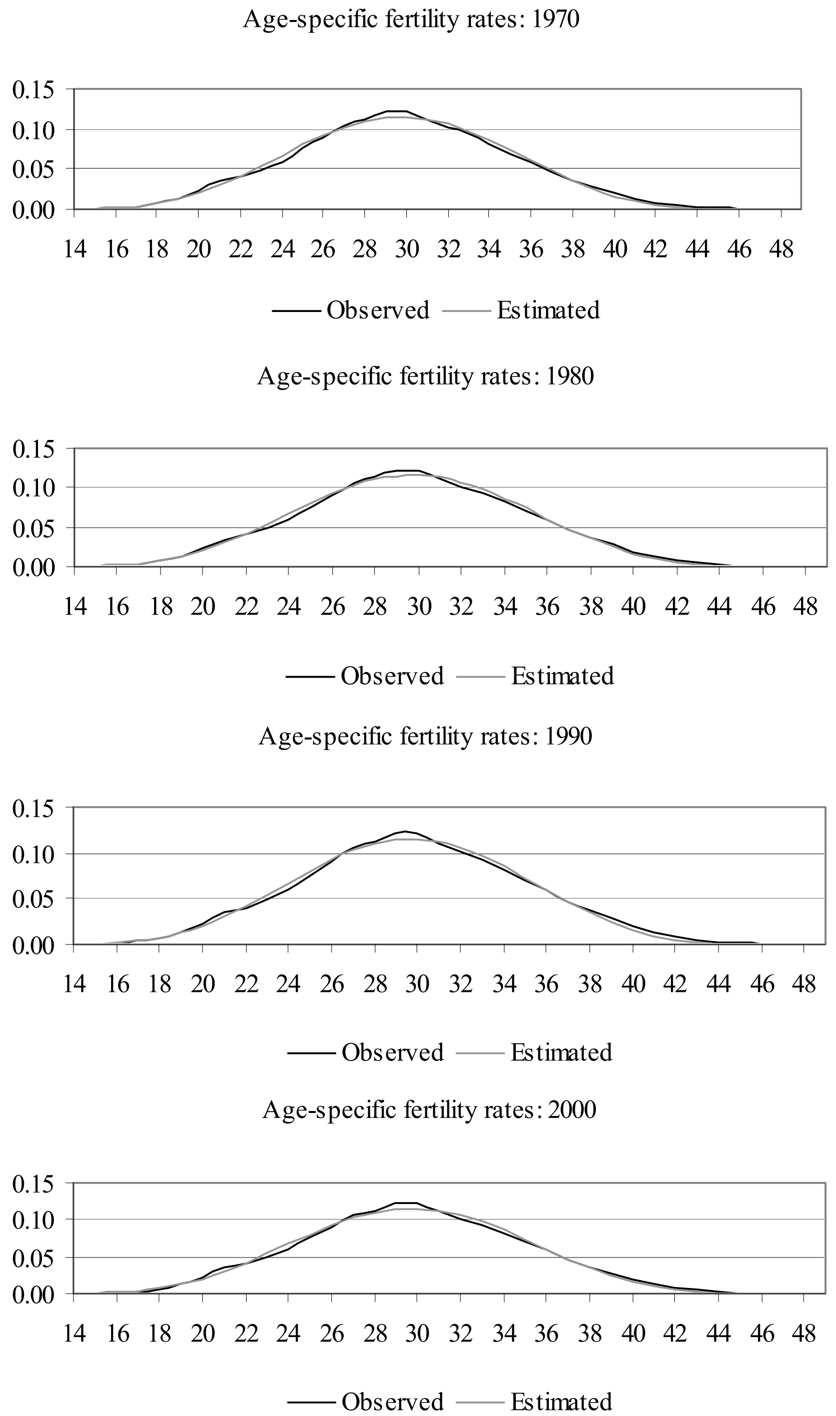
To serve as a model of a fertility schedule, we include parameters a, b and c in t(x) that control height, width and placement of the test function and obtain

$$g(x;b,c) = \begin{cases} a\,e^{\frac{-b}{1-\left(\frac{x-c}{25}\right)^2}} & \text{if} \quad -1 < \frac{x-c}{25} < 1 \\ 0 & \text{elsewhere} \end{cases} \quad . \qquad (3.1)$$

Tab.3.1. Estimation errors

Year	Number of children born				
	Estimated	Observed	Absolute error	Percent error	Estimation error
1970	110565	110149	416	0.38	0.050
1971	115293	114485	808	0.71	0.050
1972	113600	112273	1327	1.18	0.048
1973	111058	109663	1395	1.27	0.045
1974	111836	109874	1962	1.79	0.043
1975	105502	103631	1871	1.81	0.045
1976	100063	98345	1718	1.75	0.043
1977	97584	96057	1527	1.59	0.042
1978	94299	93222	1077	1.16	0.041
1979	97304	96255	1049	1.09	0.038
1980	97949	97062	887	0.91	0.037
1981	94754	94064	690	0.73	0.036
1982	93263	92748	515	0.56	0.033
1983	92079	91780	299	0.33	0.032
1984	93675	93889	214	0.23	0.031
1985	97873	98462	589	0.60	0.032
1986	101153	101950	797	0.78	0.031
1987	104310	104699	389	0.37	0.031
1988	111568	112080	512	0.46	0.029
1989	115843	116022	179	0.15	0.033
1990	123913	123934	21	0.02	0.033
1991	123716	123736	20	0.02	0.033
1992	122679	122847	168	0.14	0.031
1993	118125	117997	128	0.11	0.029
1994	112525	112257	268	0.24	0.027
1995	103991	103422	569	0.55	0.026
1996	96106	95297	809	0.85	0.020
1997	91096	90502	594	0.66	0.019
1998	89695	89028	667	0.75	0.016
1999	88748	88173	575	0.65	0.014
2000	91348	90441	907	1.00	0.016

Fig. 3.2. Observed and estimated age-specific fertility rates: Sweden, 1970-2000



The number 25 appearing in the formula is chosen by trial and error and is used to cut of the edges of the test function. We fit g(x; b, c) = g(x) to observed age-specific fertility rates $f_t(x)$ by minimizing $\sum_{x=14}^{49}(f_t(x) - g(x))^2$ with respect to its parameters and subject to the condition $\sum_{x=14}^{49} g(x) = TFR_t$ .The column labeled estimation error in Table 3.1 contains estimation errors calculated by

$\sqrt{\sum_{x=14}^{49}(f_t(x) - f^*(x))^2}$ for t = 1970, …, 2000.

These estimation errors together with the diagrams in Figure 3.2 show that estimates get more precise with time. This is largely due to the fact that the fertility curve with time becomes increasingly symmetric about its maximum value (see Figure 3.2). The estimated number of children born, as presented in Table 3.1, is calculated by applying the estimated age specific birth rates for a year to the observed mid-year population of women of that year. The errors in these estimates show a rather different pattern. For example the estimation error for the year 1970 is rather large and the estimated curve for that year is not very similar to the observed curve (fig. 3.2). On the other hand, the error in percent in the number of children born is much smaller this year then in for example the year 2000. This demonstrates that the underlying age structure can make small errors in the estimated fertility curve blow up (and at the same time make others disappear). At any rate the errors in this small experiment is smaller then those generated by the model in Section 2.

4 Parameter Estimation

In this section we look at the time development of the TFR and the parameters b and c (in our setup this is equivalent to looking at the parameters a, b and c) and we fit appropriate time series models to these. To keep the exposition focused we will not present any model diagnostics here. Application of various time series models suggests that an autoregressive model of order two provides an effective description of Swedish TFR. In choosing an appropriate model for an observed time series it is often difficult to decide how many years of data one should use in the estimation. We have after testing chosen to use TFR for the years 1970-2000 (see Figure 4.1). Using a least-squares method we have thus estimated an AR(2)-model

$$x_t = a_1 x_{t-1} + a_2 x_{t-2} + \mu(1 - a_1 - a_2) + e_t. \qquad (4.1)$$

with estimated parameters

$\hat{a}_1 = 1.56,\ \hat{a}_2 = -0.71,\ \hat{\mu} = 1.78$

The residuals (e_t) behave statistically not different from normal white noise with mean zero (and standard deviation $\hat{\sigma}_e = 0.05$), as they should.

For the parameter b an auto-regressive model of degree two also turned out to be an appropriate choice. The estimated coefficients for the model (4.1) are

$\hat{a}_1 = 1.19,\ \hat{a}_2 = -0.21,\ \hat{\mu} = 9.31.$

Fig. 4.3. Estimated and modeled parameter c

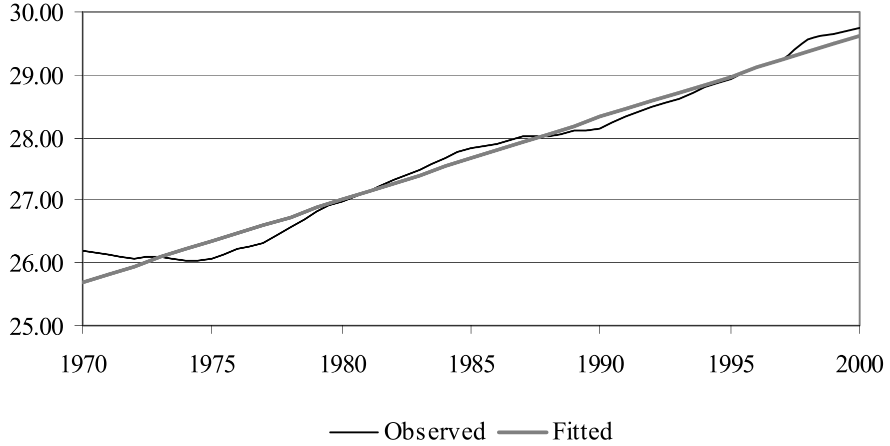
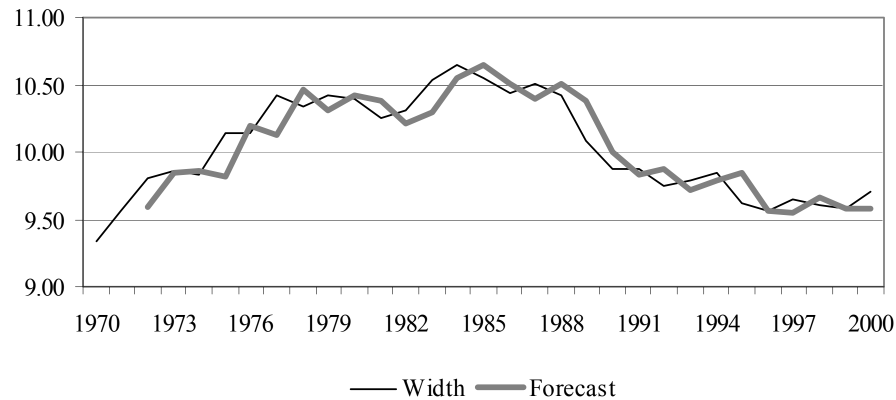


Fig. 4.2. Estimated and modeled parameter b (controlling width)



Again the residuals are (close enough) normally distributed white noise with zero mean (and with standard deviation $\hat{\sigma}_e = 0.15$). For the constant c a linear model turns out to provide the best choice. The model is

$$c = 0.13t - 232.6 + e_t$$

and the residuals have standard deviation of $\hat{\sigma}_e = 0.17$. These models can now be used for projections of TFR, b and c and the projected parameters can be put in (3.1) to give projected age specific fertility rates. We produce several projections for each parameter where we let the e_t series consist of different sets of normally distributed independent random numbers with mean zero and standard deviation as observed in past data. Figure 4.4 shows 20 projections of TFR and in Figure 4.5 we have 20 possible fertility curves for the year 2030.

Fig. 4.4. Observed and simulated TFR: 1970-2030

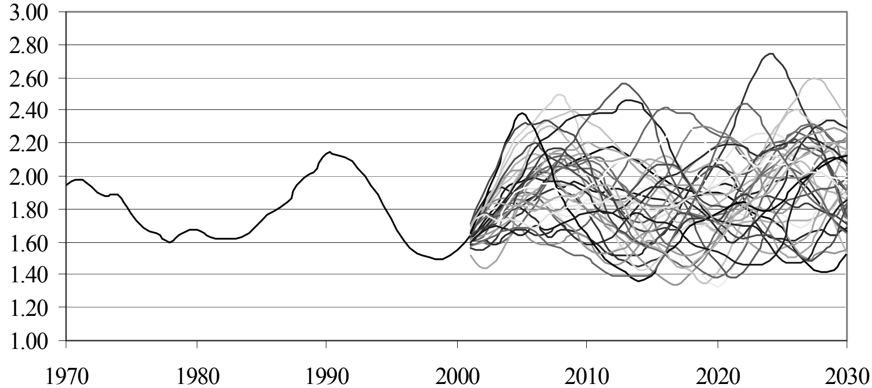
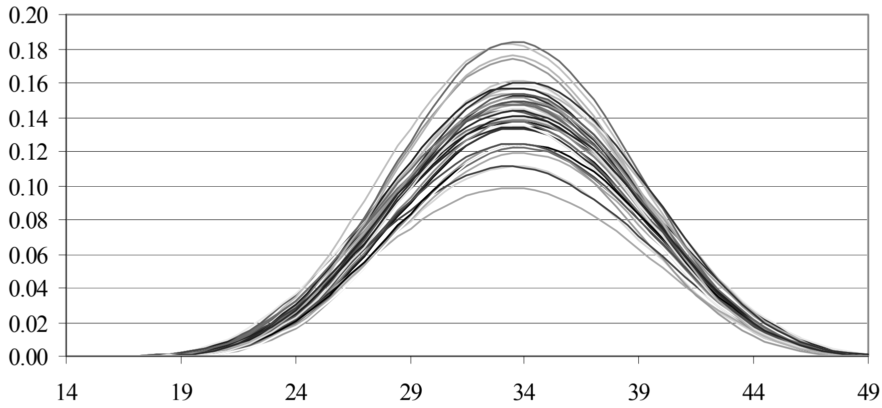


Fig. 4.5. Projected age-specific fertility rates: 2030



References

Hörmander, Lars (1963). The Analysis of Linear Partial Differential Operators I*: s Distribution Theory and Fourier Analysis.* Berlin: SpringerVerlag.

Stochastic Population Projections for Sweden

This report discusses and illustrates the thinking behind time series based stochastic population projections. Projections of this nature build on the view that demographic variables emerge from realizations of stochastic processes. For this reason the making of stochastic population projections involves representing mortality, fertility and netmigration by means of realistically chosen time series models. The advantage of this approach is that prediction intervals that reflect past variability in the demographic processes can be estimated.

The technical section in the report addresses stochastic processes in the perspective of population projections and presents three technical papers on mortality and fertility. The papers on mortality discuss the Brass logit life table survival model and how it facilitates making the survival function a random function. The paper on fertility discusses a new exponential stochastic model of age-specific fertility rates.

ISSN 1653-7149
ISBN 91-618-1346-X
ISBN 978-91-618-1346-9

Publication Services:
E-mail: publ@scb.se, phone +46 19 17 68 00, fax +46 19 17 64 44, Address: SE-701 89 Örebro.

Information and Library: E-mail: information@scb.se, phone +46 8 506 948 01, fax +46 506 948 99.
Over-the-counter sales on our premises: Statistics Sweden, Library, Karlavägen 100, Stockholm, Sweden.